Fall 9-9-2014

Selection and Genetic Drift in North American Maize

Collin Lamkey
University of Nebraska-Lincoln

SELECTION AND GENETIC DRIFT IN NORTH AMERICAN MAIZE

by

Collin M. Lamkey

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Agronomy & Horticulture

Under the Supervision of Professor Aaron J. Lorenz

Lincoln, Nebraska

September, 2014

SELECTION AND GENETIC DRIFT IN NORTH AMERICAN MAIZE

Collin M. Lamkey, Ph.D.

University of Nebraska, 2014

Advisor: Aaron J. Lorenz

Characterizing the impact of selection and genetic drift in the formation of heterotic groups and patterns in maize can reveal important insights into the mechanism underlying adaptation, and the relative importance of each force in defining population structure. The objectives were to characterize the role selection for hybrid performance had in defining population structure in both a reciprocal full-sib selection (RFS) program and a large collection of historically important inbred lines.

The Illumina GoldenGate Assay was used to genotype the University of Nebraska-Lincoln Replicated Recurrent Selection (UNL-RpRS) program. Eight cycles of S1-progeny and RFS selection were conducted for an index approximating grain yield. The distance between S1-progeny programs was compared to the distance between RFS selection programs. No evidence was found to suggest a significant genome-wide impact of selection for hybrid performance. This result suggests that, genome-wide, selection was not a strong force in diverging populations.

To further investigate the roles of selection and genetic drift a second dataset was generated with genotype-by-sequencing data accompanied by increased sample size for each population in the UNL-RpRS program. A dense physical map was generated, which allowed genomic localization of selection signatures associated with directional selection

and also selection for hybrid performance. The RFS and S1-progeny selection programs left similar signatures of selection across the genome. A scan for directional selection identified similar regions under selection across replicate populations, which suggests that adaptation is occurring from standing genetic variation.

A large collection of inbred lines was collected and grouped into four eras, which represented the double-cross, three-way cross to single-cross transition, single-cross, and advanced single-cross eras of maize breeding, respectively. A small number of inbred lines were found to contribute to the parentage of the next era. The inbred lines identified here were also major contributors in other studies as well. Scanning the genome for localized selection signatures revealed genes putatively associated with cold tolerance and resistance to fungal and bacterial pathogens, which is consistent with the notion that selection for increased yield has selected hybrids with increased tolerance to biotic and abiotic stresses.

ACKNOWLEDGEMENTS

Scientific papers generally represent a project progressing, in a very linear fashion, from point A to point B. This is a very efficient and appropriate methodology to present scientific research, however, rarely is the path from A to B a straight line. Instead, the scientific process can become convoluted with twists and turns stemming from a variety of issues. Over the course of this project, I encountered several seemingly derailing issues. Without the support of several people involved I would most likely still be in the lab wondering exactly where point B lay hidden.

Any graduate student project must start with an advisor, and I am very grateful to Dr. Aaron Lorenz for taking his time and helping me to develop this project from an idea to a dissertation. A graduate student must eat, and I am very grateful to Pioneer Hi-Bred Intl. for funding me on the Jack Cavanah Honorary Plant Breeding fellowship. I would also like to acknowledge my committee for their support throughout this process. The Lorenz Lab grew nearly exponentially during my time at UNL. Sharing papers, research progress, and ideas at lab meetings greatly increased my breadth of knowledge concerning the variety of research areas the Lorenz Lab is associated with.

My time in the turf office has not been wasted. I can now differentiate between Kentucky Blue grass and Annual Rye grass, but I think I should stick to my day job.

I would also like to thank my family. My parents have supported me and provided encouragement over the last 27 years, and without them I would not be where I am today. Finally, I would like to thank my girlfriend, Gabriel, for being crazy enough to date a plant breeding-powerlifting-meathead.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

1 CHAPTER 1 INTRODUCTION

The history of maize breeding can broadly be summarized into three eras, open-pollinated, double-cross, and single-cross. The open-pollinated era was characterized by open-pollinated landrace populations developed out of Corn Belt Dent and primarily maintained and selected by farmers. The implementation of the double-cross hybrid in the United States drastically increased maize yields over open-pollinated varieties. Single-cross hybrids, developed directly from inbred lines, brought uniformity to the farmer's field and simplified production practices. It is natural to contemplate the origins of hybrid maize breeding and wonder if any genetic changes have accompanied the drastic and consistent increases in hybrid maize yields. The remainder of the introduction discusses the origins and genetic changes in maize breeding empirically and experimentally.

Hybrid maize breeding started when George Shull (1908) realized that an open-pollinated field of maize was a complex array of cryptic hybrids. Shull's (1908) revelation has three components i) select any two individuals from an open-pollinated population, ii) inbreed these two individuals to homozygosity and, iii) cross the newly developed inbred lines to obtain a reproducible F1 hybrid. F1 hybrids offer many advantages over double-cross or three-way cross hybrids. Theoretically, F1 hybrids are easy to produce because they are developed directly from inbred lines. Additionally, F1 hybrids are genetically uniform and display increased vigor over either inbred parent.

Shull published his seminal paper, “The Composition of a Field of Maize” in 1908, however, single-cross hybrids were not widely grown until the 1960s. Clearly, it took over 50 years to put Shull’s idea into practice. What happened? Two issues confronted early maize breeders. Initially, early maize breeders had to overcome severe inbreeding depression in first cycle inbred lines. Inbreeding depression was dealt with in two ways. Jenkins (1918) developed the idea of the double-cross hybrid, which is produced from two F1 hybrids each with potentially different sets of inbred parents. Double-cross hybrids were widely grown from the late-1930s through the 1960s. Simultaneously, early maize breeders crossed inbred lines superior to their contemporaries to develop more, but also better inbred lines. This development and improvement of inbred lines, itself, gave rise to the second issue early corn breeders faced. How should inbred lines be grouped to maximize the probability of obtaining a superior hybrid?

The topic of grouping inbred lines was on the agenda of the Tenth Corn Improvement Conference of the North Central Region in 1949 (Anon, 1949). The members of the conference decided to develop a committee responsible for grouping inbred lines for breeding purposes. Interestingly, the committee decided to group inbred lines based on their numbering from the 1948 inbred uniformity trials; odd number entries went into the A group, and even numbered entries went into the B group (Anon, 1949). This was the beginning of heterotic groups in maize breeding. A definition for heterotic groups was given by Melchinger and Gumber (1998) as “…a group of related or unrelated genotypes from the same or different populations, which display similar

combining ability and heterotic response when crossed with genotypes from other genetically distinct germplasm groups." Heterotic patterns are defined as heterotic groups which combine well.

Taking a step back to examine the A and B groups of inbred lines in developing breeding families and hybrids some observations can be made. Inbred lines within the A or B groups are used to make breeding families, which in turn give rise to more but also improve inbred lines. New and improved inbred lines are then crossed between the A and B groups to develop hybrids. Superior hybrids are identified and the inbred parents are then recycled. This process of inbred and hybrid development can be termed the inbred-hybrid system of maize breeding. van Heerwaarden et al. (2012) examined genetic changes in a representative set of inbred lines and found the development of very pronounced heterotic groups (Fig 1 from van Heerwaarden et al., 2012; Figure 4.1 this manuscript). Although the origins of the heterotic groups in van Heerwaarden et al. (2012) do not coincide with the A and B heterotic groups from Anon (1949), empirically the inbred-hybrid system of maize breeding gave rise to the development of very strong population structure.

Although maize breeders were very invested in developing new inbred lines for breeding purposes, they were also interested in studying population improvement methods. Reciprocal recurrent selection (RRS) is a population improvement method that is designed to improve two populations simultaneously by selecting on the population cross (Comstock et al., 1949). That is, either a half-sib (RRS) or full-sib (reciprocal full-sib selection) family is developed by crossing two different populations together.

Reciprocal recurrent selection and its derivative reciprocal full-sib (RFS) selection very closely resemble the inbred-hybrid system of maize breeding. Examining genetic changes, which occur as a result of selection from either RRS or RFS can reveal insights into the inbred-hybrid system of maize breeding.

The Iowa Stiff Stalk Synthetic to Iowa Corn Borer Synthetic No. 1 (BSSS-BSCB1) is a rigorously studied RRS program, and was most recently evaluated by Gerke et al. (2013). Examining a principle component plot of the BSSS-BSCB1 RRS program (Figure 1 from Gerke et al., 2013) reveals a very similar development in population structure to that observed empirically in the inbred-hybrid system of maize breeding. Broadly, two trends are observed both empirically and experimentally: i) the development of strong population structure with time accompanied by ii) a reduction in genetic diversity with time.

In quantitative genetics two forces are responsible for changing allele frequencies. Systematic forces, like selection, are predictable in amount and directional. Dispersive forces, like genetic drift, are random and predictable only in amount. Assuming an infinitely large base population becomes subdivided, and mating is restricted to within subpopulations, quantitative genetics theory tells us that variation will decrease within subpopulations and variation will increase among populations with time (Falconer and Mackay, 1996). Overtime, a very similar trend in the development of population structure is observed compared to either the inbred-hybrid system of maize breeding or RRS.

The critical question this dissertation tries to answer is whether the power of selection to change allele frequencies is greater than the power of genetic drift. In other words, is the observed genetic structure, both empirically and experimentally, the result of selection for hybrid performance or genetic drift? The structure of the remaining chapters is similar to the introduction, except I evaluate the effect of selection for hybrid performance experimentally in chapters two and three and empirically in chapter four.

The first objective of this dissertation was to evaluate the effect of selection for hybrid performance in diverging base populations within a recurrent selection program. In chapter 2, I evaluate this objective by using genome-wide estimates of genetic distance in two different population improvement methods of a recurrent selection program developed at the University of Nebraska-Lincoln. The second objective was to identify regions of the genome which have likely been under directional selection after eight cycles of selection. In chapter 3, I evaluate this objective in addition to a search for selection for hybrid performance across the genome within the same recurrent selection program. The final objectives were to describe the genetic structure in a large collection of inbred lines and search for putative selection candidates. In the fourth chapter I evaluate this objective by assembling a large collection of historically important inbred lines. I conclude the dissertation by drawing overall conclusions from chapters two, three, and four.

1.1 References

Anon. 1949. Minutes of the meeting of the North Central Corn Breeding Technical Committee. USDA. Beltsville, Maryland.

Comstock, R.E., H. Robinson, and P. Harvey. 1949. A breeding procedure designed to make maximum use of both general and specific combining ability. Agron. J. 41:360-367.

Falconer, D.S., and T.F.C. Mackay. 1996. An introduction to quantitative genetics. Pearson Prentice Hall, Essex, England.

Gerke, J.P., J.W. Edwards, K.E. Guill, J. Ross-Ibarra, and M.D. McMullen. 2013. The genomic impacts of drift and selection for hybrid performance in maize. ArXiv Preprint arXiv:1307.7313.

Melchinger, A.E., and R.K. Gumber. 1998. Overview of heterosis and heterotic groups in agronomic crops. Concepts and Breeding of Heterosis in Crop Plants29-44.

Shull, G.H. 1908. The composition of a field of maize. J. Hered.296-301.

van Heerwaarden, J., M.B. Hufford, and J. Ross-Ibarra. 2012. Historical genomics of north american maize. Proc. Natl. Acad. Sci. U. S. A. 109:12420-12425.

2 CHAPTER 2 RELATIVE EFFECT OF DRIFT AND SELECTION IN DIVERGING POPULATIONS WITHIN A RECIPROCAL RECURRENT SELECTION PROGRAM

2.1 Abstract

The establishment of a solid heterotic pattern has been beneficial to maize (*Zea mays* L.) breeding in North America. It has been shown that genetic divergence between heterotic groups has increased over time, but the relative importance of drift versus selection for hybrid performance on the increased divergence is not known. Reciprocal recurrent selection is a systematic method of selection on hybrid performance, where two populations are selected and improved on the basis of the population-cross performance. The University of Nebraska carried out a replicated recurrent selection program (UNL-RpRS) for eight cycles of selection. Three replicate programs were conducted in parallel for both reciprocal full-sib selection (RFS) and S1-progeny selection. The objective of this study was to evaluate the effect of selection for hybrid performance on population divergence using the UNL-RpRS as a model system. Genome-wide divergence among the base populations increased over eight cycles of selection. Divergence was not, however, greater among base populations subjected to RFS compared to S1-progeny selection. Moreover, F_{st} values among replicate populations within a base population and selection method were as large as F_{st} values between complementary populations selected for population cross performance using RFS. A small increase in divergence through RFS relative to S1-progeny selection was observed when markers were filtered for high levels

of adjacent linkage disequilibrium, but the effect was consistent across replicates. We did not consistently detect effects of selection for hybrid performance on genome-wide divergence between populations, suggesting selection was not a strong force in diverging populations in the UNL-RpRS.

2.2 Introduction

Hybrid maize breeding entails identifying two parents that, when crossed together, produce a hybrid of superior performance through both additive effects and heterosis. Testing many combinations of candidate parents greatly complicates the breeding pipeline. According to Tracy and Chandler (2006), in the early days of hybrid maize breeding it was not entirely clear which sets of inbred parents produced the best hybrids, although there was a sense that genetic distance between lines maximized heterosis. To organize maize breeding programs and make them more efficient, it was suggested at the 1949 North Central Regional Corn Improvement Conference that existing inbred lines be sorted into two groups so as to avoid relationships between inbred lines used as parents of hybrids (Tracy and Chandler, 2006). Over time it was recognized that Iowa Stiff Stalk Synthetic (BSSS) germplasm performed well as females and lines that combined well with BSSS germplasm, termed non-Stiff Stalk, were eventually sorted into a male group (Tracy and Chandler, 2006). Selection and breeding proceeded for many decades by advancing parent lines based on their performance when crossed to parent lines from the opposite heterotic group.

Very little population structure existed between the designated female and male groups during the early years of hybrid maize breeding. Genetic divergence between

groups has occurred over time and has largely been driven by breeders. This effect has been documented in earlier studies using microsatellite markers (Duvick et al., 2004) and restriction fragment length polymorphisms (Hagdorn et al., 2003). More recently, van Heerwaarden et al. (2012) genotyped 400 maize accessions, ranging from landraces to elite inbred lines using the Illumina MaizeSNP50 beadchip. This study nicely showed that population structure has greatly increased over time due to the decreasing number of effective ancestors contributing to the genetic diversity of later generations of maize inbred lines used by the North American seed industry.

Recurrent selection programs targeting inter-population improvement have shown similar signs of increased divergence between complementary populations. Recurrent selection is a cyclical procedure designed to gradually improve the mean performance of a population over time by increasing the frequency of favorable alleles while maintaining genetic variation to allow continual genetic gain (Hallauer and Miranda, 1988). At the most general level, two types of recurrent selection can be used: intra-population and inter-population improvement. Intra-population improvement methods are designed to improve the performance of the population *per se*, while inter-population improvement methods are designed to increase the performance of the population cross (Fehr, 1991). Considerable debate about the model of dominance in maize prompted the creation of reciprocal recurrent selection (RRS), an inter-population improvement procedure proposed by Comstock et al. (1949) to improve the population cross despite the mode of gene action.

The Iowa Stiff Stalk Synthetic (BSSS) and Iowa Corn Borer Synthetic (BSCB1) RRS program was initiated in 1949 (Penny and Eberhart, 1971) and has completed 17 cycles of selection (Brekke et al., 2011). Molecular characterization of these populations using microsatellites and RFLPs has revealed that genetic distance between the populations has increased over cycles, with accompanying decreases in genetic variance within populations (Labate et al., 1997; Hinze et al., 2005; Gerke et al., 2013). Similar observations have been made by molecular characterization of RRS in Spanish populations (Romay et al., 2012), tropical populations (Solomon et al., 2010), and other temperate populations (Butruille et al., 2004).

The above results clearly indicate heterotic groups diverge over time. What is not so clear, however, is the predominant force underlying this divergence. Assuming an over-dominance or pseudo-overdominance model of heterosis, we would expect selection for hybrid performance between two populations to drive frequencies of complementary alleles in each population in opposite directions, thereby maximizing the probability of obtaining heterozygotes in the hybrids. Genetic divergence between isolated groups is also expected through pure random genetic drift (Falconer and Mackey, 1996). None of the aforementioned studies were able to determine the relative importance of selection versus genetic drift in the divergence of heterotic groups accompanying decades of breeding for hybrid performance in maize.

The University of Nebraska Replicated Recurrent Selection Program (UNL-RpRS) was initiated in 1968 to compare two methods of recurrent selection: S1-progeny selection (intra-population improvement) and reciprocal full-sib selection (RFS; inter-

population improvement). A highly unique characteristic of this recurrent selection program was the use of replicated selection. Each recurrent selection method was independently replicated three times within each base population, allowing the originator, Professor William Compton, to separate the effects of selection and genetic drift for a more precise comparison. Eight cycles of selection for grain yield and lodging were conducted for each method and replicate. A full review of the UNL-RpRS is provided in the Materials and Methods section.

Populations derived from the UNL-RpRS can be used to study several aspects related to selection in maize breeding, one of which is the role of selection in diverging heterotic groups. Illuminating the role of selection and drift in the establishment of strong heterotic patterns would be informative for the enhancement of current heterotic patterns, as well as the creation of new heterotic patterns in nascent hybrid breeding programs. The objective of this study was to evaluate the effect of selection for hybrid performance on population divergence within the UNL-RpRS.

2.3 Materials and Methods

2.3.1 Development of the UNL Replicated Recurrent Selection Populations

Three populations composed the germplasm used in the UNL-RpRS: Nebraska Krug (NK), Nebraska Stiff Stalk Synthetic, and Nebraska B Synthetic. Nebraska Krug is an open-pollinated variety commonly grown in eastern Nebraska before the development of hybrid maize (Thomas, 1979). Nebraska Stiff Stalk Synthetic (NSS) was derived from BSSS in 1948 after two cycles of selection for general combining ability, and

subsequently grown and adapted to Nebraska (Odhiambo, 1987). Nebraska B Synthetic (NBS) was created by Lonnquist in 1946 by intercrossing 32 inbred lines of diverse origin (Supplementary Table 1).

The UNL-RpRS commenced in 1968 by sampling the first replicate populations from NK, Neb SS, and NBS. The S1-progeny selection and RFS programs were started from independent samples of individuals (Figure 1). Reciprocal recurrent selection proceeded by making 100 full-sib families from NBS x NK and NBS x NSS. S1 families were simultaneously made by using two-eared plants. The second ear shoot was self-pollinated, followed by the top most ear shoot being cross-pollinated by a plant from the reciprocal population. Additional details of the crossing procedure are provided in Odhiambo (1987). Full-sib families created in 1968 were tested in 1969 in unreplicated trials at three irrigated Nebraska locations (West, 1980). Selections were based on average performance for the multiplicative index (Elston, 1963)

$$I = Y(1-L)(1-D)$$

where *Y* represents yield adjusted to 15.5% moisture, *L* represents proportion of lodged plants, and *D* represents proportion of dropped ears. Ten full-sib families were selected, and the corresponding S1-progeny seed from each parent of selected full-sib families within each population were recombined. Chain crossing was used to ensure each S1 family was crossed to at least four other S1 families.

In the RFS program selection was based on full-sib families and recombination was based on S1 families. S1 family (S1) selection is an intra-population improvement

procedure where selection and recombination were based on S1 families. At harvest, equal amounts of seed were taken from each cross to form a balanced bulk comprising replicate one of cycle one. Replicates two and three were initiated in 1969 and 1970, respectively, following the same procedures.

The first replicate of S1-progeny selection was also initiated in 1968 by independently sampling individuals from the base populations NKS, NSS, and NBS. Briefly, 100 S_1 families were created selecting plants with at least two ear shoots. Prior to silking, the top two ear shoots were covered with glassine bags. At anthesis, the second shoot was self-pollinated and top most shoot was de-bagged and allowed to open pollinate. At harvest, only those plants where seed was set on both ears were selected to form the 100 S1 families. This procedure was carried out to impose the same selection for prolificacy as that in the RFS program (Odhiambo, 1989). S_1 families were evaluated under irrigated and non-irrigated conditions in Lincoln, NE using a RCBD with two replications. Ten S_1 families from each population were selected on the basis of the same multiplicative index used in the RFS program. Selected families were recombined and bulked using the same methods as the RFS program. Replicates two and three of the S1-progeny selection program were initiated in 1969 and 1970 just as in the RFS program. Therefore, fifteen parallel selection programs were being managed: three replicates of RFS in each of NBS x NKS and NBS x NSS, and three replicate S1-progeny selection programs in each of the NKS, NSS, and NBS populations. Replicates were staggered across years for logistical reasons.

The selection programs involving NKS were dropped after cycle five (Dana Galusha, personal communication). After cycle five, hand planting of hill plots and hand harvesting were replaced by machine planting and harvesting. Eight cycles of selection were completed for each replicate and recurrent selection method.

2.3.2 Genotyping

Nineteen or twenty individuals were sampled from each of the 14 populations for genetic analysis (Table 1). Genotyping was performed using an Illumina GoldenGate Assay on 768 SNPs (Jones et al., 2009). Forty-one markers failed outright. Markers were retained for further analysis if minor allele frequency was greater 10% and missing data frequency was less than 10%. The final dataset contained 274 individuals and 513 markers. Missing marker scores were imputed using the Beagle (Browning and Browning, 2009) software implemented in the R package Synbreed (Wimmer et al., 2012).

2.3.3 Statistical Analysis

Rate of linkage disequilibrium (LD) decay was assessed by plotting r^2 between markers against physical distance. PowerMarker V3.25 (Liu and Muse, 2005) was used to compute population summary statistics. Observed heterozygosity (Ho), expected heterozygosity (D), and number of polymorphic markers were chosen to summarize within population genetic diversity. Observed heterozygosity and D were estimated as described by Weir (1996) and averaged over all loci within a subpopulation. Average estimates were obtained by averaging over all loci within a subpopulation. Bootstrap estimates of standard errors for observed and expected heterozygosity were obtained by

resampling markers 10,000 times. Gene diversity is defined as the probability that two randomly chosen alleles from the population are different, and should be close in value to observed heterozygosity in outbred populations (Weir, 1996). The unbiased option was selected for computation of gene diversity which weights the estimate by the within population inbreeding coefficient. Allele number is the average number of alleles in each population. Effective population size (Ne) was computed by rearranging equation 3.13.3 from Kimura and Crow (1970):

$$N_e = \frac{1}{2(1-\sqrt[t]{P})}$$

where $P = \frac{H_t}{H_0}$ is the panmictic index, t is time in generations, H_t is the observed heterozygosity in cycle 8, and H_o is the observed heterozygosity in cycle 0 (Table 2).

Modified Roger's Distance (MRD) is a scaled Euclidean distance computed as:

$$MRD = \frac{1}{\sqrt{2m}}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(p_{ij}-q_{ij})^2}$$

where p_{ij} and q_{ij} are the allele frequencies of the j^{th} allele at the i^{th} locus between two populations. The standard error for each element of the MRD matrix can be obtained with a jackknife estimator by dropping a locus and computing the MRD (Efron and Tibshirani, 1986; Hagdorn, 2003). For each element of the distance matrix, the l^{th} locus was omitted and the MRD was computed. This was repeated m times for each element of the MRD matrix. The variance was found as:

$$V(MRD) = \frac{m-1}{m}\sum_{j=1}^{m}(MRD_{\_j} - MRD_m)^2$$

where m represents the number of loci, $MRD_{\_j}$ represents the MRD with the j^{th} locus omitted, and MRD_m represents the mean of m estimates of $MRD_{\_j}$ (Efron and Tibshirani, 1986; Hagdorn et al., 2003).

Analysis of Molecular Variance was performed using Arlequin v3.0 (Excoffier et al., 2005) to quantify variation among and within sets of populations. Two scenarios were considered. First, random genetic drift can be quantified by grouping the three replicate populations within a base population and breeding method. Second, variance among and within each cycle can be quantified by comparing each C0 and C8 combination. Significance of among population variation and F_{st} was established by permuting individual genotypes among populations 1000 times (Excoffier et al., 2005).

Principal component analysis (PCA) was conducted by first recoding the marker matrix to 0, 0.5, and 1 corresponding to the number of major alleles present in the original genotype matrix. The correlation among individuals was computed based on the recoded marker matrix, followed by eigen analysis on the correlation matrix.

2.4 Results

2.4.1 Review of Genetic Gain results reported for UNL-RpRS

Results on the response to selection and its variation within the UNL-RpRS are reviewed here in order to provide some background and context to the population genetic analysis performed for the study reported herein. Because the NKS programs were

discontinued, discussion of results will be restricted to NBS and NSS. Significant genetic gain was achieved through selection for performance of the population cross and S1-progeny performance per se (Figure 2; Tragresser et al., 1989). Genetic gain for the performance of the randomly mated population was achieved in NBS, but not NSS (Odhiambo and Compton, 1989; Galusha, 1999). Panmictic mid-parent heterosis (PMPH) is the difference between an F1 hybrid and the mean of the two panmictic populations from which it was formed (Lamkey and Edwards, 1999). After five cycles of selection PMPH was 20 % and 16 % for S1 and RFS selection, respectively (Odhiambo, 1987), but after eight cycles of selection PMPH (for yield) was 26 % and 42 % for S1 and RFS selection, respectively, indicating RFS was superior for increasing heterosis. This result is mostly explained by the observation that C8 yields of the NBS and NSS populations per se were higher for the S1-progeny selection program than for the RFS selection program (Galusha, 1999). Based on the fact that population per se performance was improved by S1-progeny selection, it appears S1-progeny selection was more successful at exploiting additive genetic variation, while RFS must have tapped into dominant, overdominant, or pseudo-overdominant gene action to improve the population cross.

Inbreeding depression is the reduction in population performance due to an increased inbreeding coefficient of individuals (Falconer and Mackay, 1996), and can be measured as the decrease in the population mean after one generation of self-fertilization compared to the respective random mated populations (Thomas, 1979). After two cycles of S1-progeny selection, inbreeding depression was measured to be 19 and 33 % for the NBS and NSS populations, respectively (Thomas, 1979). After two cycles of RFS

selection, inbreeding depression was 28 and 42 % for the NBS and NSS populations, respectively, suggesting a large advantage of S1-progeny selection for reducing inbreeding depression possibly through purging deleterious alleles. Galusha (1999), on the other hand, found the levels of inbreeding depression to be more similar between the selection methods, being only four and five percentage points different for NBS and NSS, respectively.

On average, RFS and S1-progeny selection performed similarly with respect to improvement of the population cross (West et al., 1980; Odhiambo, 1987; Galusha, 1999). S1-progeny selection has been recommended because it considerably improved averaged inbred-progeny performance (Galusha, 1999). Considerable variation among the replicates, however, has been reported and thwarts comparisons among methods (Figure 2). This variation is caused by both random genetic drift (Tragesser et al., 1989) and selection environment effects. Galusha (1999) emphasized the importance of founder effects that drove the trajectory of population improvement for each replicate selection program. It can be seen in Figure 2 that the population cross performance of replicate three was reported to be considerably greater in both cycle five and cycle eight compared to the other replicates within both the RFS and S1-progeny selection programs.

2.4.2 Linkage Disequilibrium Decay

Number of polymorphic markers per subpopulation ranged from 217 to 459 (Table 2). Distribution of r^2 values between adjacent markers is displayed in Supplementary Figures 1 and 2. On average, LD between adjacent markers was found to be low, with a median value of 0.14. Markers were not well distributed over

chromosomes and generally were concentrated near ends of chromosome arms (Supplementary Figure 3), possibly limiting power to detect selection signatures, especially near pericentromeric regions. On the other hand, the number of markers deployed for this study was considerably more than used in other similar studies (Hinze et al., 2005; Butruille et al., 2004; Romay et al., 2012), and 44 % of adjacent markers did have LD values greater than 0.20 (Figure LD.B).

2.4.3 Genetic variation between populations

Eight cycles of selection and recombination of 10 selected S1 families reduced observed and expected heterozygosity by 30% and 35% for NSS and NBS, respectively (Table 2). Small amounts of variation for Ho and D were observed between replicate populations. Retrospective estimates of effective population size, ranging from 7 to 13, are in line with the actual number selected. As discussed above, S1-progeny selection was clearly successful for improving average S1 performance. If S1-progeny selection were successfully purging deleterious alleles through selection on partially inbred families, we would expect the genetic variation to have been reduced more in these populations compared to the populations resulting from RFS. However, no consistent difference in Ho and D between populations that had undergone RFS and S1-progeny selection was observed (Table 2).

Variation among replicated populations can be used to estimate the degree of genetic drift in a selection program. Single locus theory predicts that the cumulative effects of random genetic drift will produce an inbreeding coefficient of 34% assuming $N_e = 10$, $t = 8$, and random mating (Falconer and Mackay, 1996). An analysis of

molecular variance was performed by grouping replicates by base population and breeding method. The among-replicate source of variation is an empirical measure of the level of inbreeding that has accumulated after eight cycles of selection. Partitioning of the RFS replicates indicated that 31.85 % (NBS) and 35.47 % (NSS) of the variation was distributed among populations. Likewise, partitioning variation in the S1 replicates indicated that 40.86 % (NBS) and 34.31 % (NSS) of the variation was distributed among replicates. With the exception of NBS replicates that had undergone S1 selection, the observed F_{st} values among replicates were close to theoretical expectations.

As expected, the genetic distance between NBS and NSS for all pairs of C8 populations was greater than the C0 populations (Table 3; Figure 3). Distances between NBS and NSS for the RFS selection and S1-progeny selection were similar (Table 3; Figure 3). One exception was the distance between replicate three of the RFS program (i.e., NBS_RFS8_3 and NSS_RFS8_3), which was slightly higher (~5%) than the next largest distance between RFS population pairs. It is noteworthy that the cross between the third RFS replicate populations also was the highest yielding by a substantial margin (Figure 2. The MRD values, however, were within the distribution of distances between all pairwise combinations of S1-progeny populations (MRD). A visual evaluation of population structure using a PCA also fails to reveal increased divergence as a result of RFS compared to S1-progeny selection (Figure 4).

Rapid LD decay within the UNL-RpRS populations suggests that a large fraction of the assayed SNPs could be selectively neutral. To investigate the effect putative selectively neutral SNPs are having on the estimated genetic distances between

populations, SNPs were filtered according to the level of LD between adjacent loci. The retained SNPs had an $r^2 > 0.20$ with at least one adjacent locus, increasing the likelihood that these SNPs are also in LD with yield QTL within the marker intervals. The number of SNPs meeting this criterion ranged from 98 to 217 across populations (Supplementary Table 2). Upon removal of putative neutral markers, it was observed that both MRD and F_{st} increased between the complementary RFS populations, while the distances remained the same between the corresponding S1 populations. This was especially the case for NBS_RFS_C8_3 and NSS_RFS_C8_3, which, as discussed above, displayed the most genetic gain under reciprocal full-sib recurrent selection for the population cross. The Fst value between NBS_RFS_C8_3 and NSS_RFS_C8_3 was 12% higher than it was between the corresponding S1 populations, NBS_S1_C8_3 and NSS_S1_C8_3. The MRD between these populations was 0.08 higher than between the corresponding S1 populations, well beyond the bootstrap-estimated standard errors. Nevertheless, the MRD of 0.52 between these populations is still less than the MRD between a pair of S1 populations, NBS_S1_C8_2 and NSS_S1_C8_1 (MRD=0.53; Table 4).

2.5 Discussion

Heterosis is most commonly expressed as the dominance effect multiplied by the allele frequency difference (Falconer and Mackay, 1996). Allele frequency differences between two operational taxonomic units (e.g., inbred lines, synthetic populations) quantified by the MRD (Melchinger, 1999), and panmictic mid-parent heterosis (PMPH; Lamkey and Edwards, 1999) is equal to the square of the MRD multiplied by the dominance effect. Under simple single-locus theory and assuming dominance effects are

in the overdominance range (caused by true overdominance or pseudo-overdominance), we would expect selection on hybrid performance to select for complementary alleles in populations, and thus diverge populations more than we would expect by genetic drift alone.

The populations developed by the UNL-RpRS provide a powerful system to test the effectiveness of RFS in diversifying populations. The three pairs of replicate populations that had undergone S1- progeny selection may serve as a null hypothesis in terms of the amount of genetic divergence we can expect by drift alone with respect to the population cross. If the pairs of complementary RFS populations had diverged from one another more than the pairs of S1-progeny populations, we could infer that selection for hybrid performance played a key role in population divergence. We did not find consistent and strong effects of selection for hybrid performance on genome-wide divergence between populations. These results suggest selection was not a major force in diverging populations within the UNL-RpRS, highlighting the potential role of drift in establishing and diverging heterotic groups. It was found that the pair of complementary populations diverged the most (RFS replicate 3) reportedly had substantially higher yield by the eighth cycle of selection compared to the other two RFS replicates, but it is not possible to clearly differentiate effects of selection from drift in the case of one population. Performance comparisons among previous studies (Thomas, 1979; Odhiambo, 1987; Galusha, 1999) found that S1-progeny selection was more successful in exploiting additive genetic variance as indicated by its greater improvement of population per se performance as well as performance after one generation of self-fertilization.

Because RFS improved the population cross slightly more than S1-progeny selection, it was suggested that RFS must be exploiting a different set of loci or genetic system, presumably composed more of gene effects in the dominance to overdominance range. The population genetic parameters calculated for this study (F_{st} and MRD) do not indicate a predominant effect in the form of overdominance, pseudo-overdominance, or epistasis leading to population divergence through selection. Similar results were found upon examination of selection signatures left by 16 cycles of reciprocal recurrent selection on the BSSS and BSCB1 populations (Gerke et al., 2013). Gerke and co-workers used computer simulations to show that most of the genetic divergence of these populations through time could be caused by genetic drift, with very little divergence left to explain by selection on the population cross despite dramatic increases in hybrid yield.

It is tempting to tout the positive effects of genetic drift on establishing heterotic groups and maximizing heterosis, but the role of genetic drift in the establishment of heterotic patterns is not that it maximizes heterosis, but rather that it increases the inbreeding coefficient of within-heterotic group crosses, ensuring that breeders are continually forced to make superior hybrids by crossing between groups. As reviewed by Tracy and Chandler (2006), the original primary purpose of the heterotic pattern was to help breeders narrow the field of potential crosses and make their programs more efficient. Connecting the results reported herein for the UNL-RpRS to the relative role of drift and selection on the divergence of heterotic groups within the U.S. seed industry (van Heerwaarden et al., 2012) would be highly speculative given differences in yield gains and selection intensities between the two systems. Nevertheless, the influence of

genetic drift was clearly displayed in the UNL-RpRS, and it would be hard to argue that drift didn't also play an important role shaping population structure of North American maize.

Founder effects can have a profound influence in the trajectory of population improvement. Several studies using replicated selection have shown that replicates will arrive at the same selection limit, but the rate at which the limit is approached could differ (e.g., Gall, 1971; Falconer, 1973). The UNL-RpRS clearly shows that some replicates had responded to selection more than others, presumably through the influence of sampling effects which established initial levels of genetic variation, although genotype-by-environment effects across the eight cycles of selection cannot be ruled out. Undoubtedly, founder effects and genetic drift greatly influenced the trajectory of genetic gain in North American maize, just as happenstance and genetic drift shaped the evolution of biological diversity in general (Gould, 1989).

Although a genome-wide signature of selection was not consistently observed across replicates, it is entirely possible that specific loci or genomic regions have diverged through RFS. For example, the potential importance of the Hill-Robertson effect on heterosis caused by repulsion phase linkages in regions of low recombination (McMullen et al., 2009) could cause alleles in the pericentromeric regions to diverge through selection on pseudo-overdominance gene action. Marker densities, particularly in pericentromeric regions, used in this study were not high enough to reveal patterns of divergence across specific genomic regions. Future studies aimed at detection of selection at this level will use higher marker densities. Another potential problem with the SNP

array used in this study is ascertainment bias. Curiously, gene diversity measures, number of polymorphic loci (Table 2) and visual inspection of the PCA plot (Figure 4) suggest that NBS is less diverse than NSS. This goes against our expectations based on the composition of the founders (Supplementary Table 1). One explanation is that the SNPs assayed for this study did not capture genetic variation contained within NBS well possibly because the founders of NBS are older and obscure inbred lines. On the other hand, Jones et al. (2009) show ascertainment bias of this genotyping assay to be minimal. Sixty diverse public inbred lines were used to select the 768 SNPs.

2.6 Conclusions

Results of this study show the importance of genetic drift in differentiating populations undergoing reciprocal recurrent selection. Despite substantial genetic gain for heterosis from reciprocal recurrent selection, we did not find consistent and strong effects of selection for hybrid performance on population divergence on a genome-wide basis. Founder effects and genetic drift likely underlie the variation in response to selection between replicates.

2.7 References

Brekke, B, J. Edwards, and A. Knapp. 2011. Selection and adaptation to high plant density in the Iowa stiff stalk synthetic maize (L.) population. Crop Sci. 51:1965-1972.

Browning, B.L. and S.R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. The American Journal of Human Genetics. 84:210-223.

Butruille, D.V., H.D. Silva, S.M. Kaeppler, and J.G. Coors. 2004. Response to selection and genetic drift in three populations derived from the golden glow maize population. Crop Sci. 44:1527-1534.

Comstock, R.E., H.F. Robinson, and P.H. Harvey. 1949. A breeding procedure designed to make maximum use of both general and specific combining ability. J. Am. Soc. Agron. 41:360-367.

Crow, J.F. and M. Kimura. 1970. An introduction to population genetics theory. Blackburn Press, Cadwell, New Jersery.

Duvick, D.N., J.S.C. Smith, and M. Cooper. 2004. Long-term selection in a commercial hybrid maize breeding program. p. 109-152. In: Jules Janick (ed.) Plant breeding reviews, part 2. John Wiley and Sons, Inc, Hoboken, New Jersey.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science. 1:54-77.

Elston, R.C. 1963. A weight-free index for the purpose of ranking or selection with respect to several traits at a time. Biometrics 19:85-97.

Excoffier, L. G. Laval, and S. Schneider (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50.

Falconer, D.S. 1973. Replicated selection for body weight in mice. Genetical Research. 22:291-321.

Falconer, D.S., and T.F.C. Mackay. 1996. An introduction to quantitative genetics. Pearson Prentice Hall, Essex, England.

Fehr, W.R. 1991. Principles of Cultivar Development. Iowa State University, Ames, IA.

Gall, G.A.E., 1971. Replicated selection for 21-day pupa weight of *Tribolium castaneum*. Theor. Appl. Gen. 41:164-173.

Galusha, D.D. 1999. An evaluation of eight cycles of replicated recurrent selection in maize. Ph.D. diss. Univ. of Nebr., Lincoln, NE.

Gerke, J.P., J.W. Edwards, K.E. Guill, J. Ross-Ibarra, and M.D. McMullen. 2013. The genomic impacts of drift and selection for hybrid performance in maize. Genetics (*In press*).

Gould, S.J. 1989. Wonderful Life. Norton and Co., New York, NY.

Hagdorn, S., K.R. Lamkey, M. Frisch, P.E.O. Guimaraes, A.E. Melchinger. 2003. Molecular genetic diversity among progenitors and derived elite lines of BSSS and BSCB1 maize populations. Crop Sci. 43:474-482.Hallauer, A.R. and J.B.

Miranda Fo. 1988. Quantitative genetics in maize breeding. Iowa State University Press, Ames, IA.

Hinze, L.L., S. Kresovich, J.D. Nason, K.R. Lamkey. 2005. Population genetic diversity in a maize reciprocal recurrent selection program. Crop Sci. 45:2435-2442.

Jones, E., W. Chu, M. Ayele, J. Ho, E. Bruggeman, K. Yourstone, A. Rafalski, O.S. Smith, M.D. McMullen, C. Bezawada, J. Warren, J. Babayev, S. Basu, and S. Smith. 2009. Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (Zea mays L.) germplasm. Mol Breeding. 24:165-176.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1997. Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. 37:416-423.

Lamkey, K.R., and J. Edwards. 1999. Quantitative genetic of heterosis. p. 31-48. In: J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Liu, K. and S.V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 21:2128-2129.

McMullen., M.D., S. Kresovich, H.S. Villeda, et al. 2009. Genetic properties of the maize nested association mapping population. Science 325:737-740.

Melchinger, A.E. 1999. Genetic diversity and heterosis. p. 99-118. In: J.G. Coors and S. Pandey (ed.) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Odhiambo, M.O. 1987. A comparison of replicated S1 line per se and reciprocal full-sib index selection in maize (*Zea mays L.*). Ph.D. diss. Univ. of Nebr., Lincoln, NE.

Odhiambo, M.O. and W.A. Compton. 1989. Five cycles of replicated S1 vs reciprocal full-sib index selection in maize. Crop Sci. 29:314-319.

Penny, L.H. and S.A. Eberhart. 1971. Twenty years of reciprocal recurrent selection with two synthetic varieties of maize (Zea mays L.). Crop Sci. 11:900-903.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Romay, M.C., A. Butron, A. Ordas, P. Revilla, B. Ordas. 2012. Effect of recurrent selection on the genetic structure of two broad-based Spanish maize populations. Crop Sci. 52:1493-1502.

Solomon, K.F., I. Martin, A. Zeppa. 2010. Temporal genetic structure patterns in tropical maize populations under reciprocal recurrent selection. Euphytica. 176:239-249.

Thomas, M. 1979. Replicated S1 line per se and reciprocal full-sib selection compared in maize (*Zea mays L.*). Ph.D. diss. Univ. of Nebr., Lincoln (Diss. Abstr. 40-4595).

Tracy, W.F., M.A Chandler. 2006. The historical and biological basis of the concept of heterotic patterns in corn belt dent maize. p. 219-233. In: K.R. Lamkey and M.

Lee (ed.) Plant Breeding: The Arnel R. Hallauer International Symposium. Blackwell Publishing, Ames, IA.

Tragesser, S.L., W.C. Youngquist, O.S. Smith, W.A. Compton. 1989. Drift vs. selection effects from five recurrent selection programs in maize. Maydica. 34:23-32.

Troyer, A.F. 2004. Background of U.S. Hybrid Corn II: Breeding, Climate, and Food. Crop Sci. 44:370 – 380. Weir, B. 1996. Genetic data analysis. Sinauer Associates Inc., Sunderland, MA.

van Heerwaarden, J., M.B. Hufford, J. Ross-Ibarra. 2012. Historical genomics of North American maize. PNAS. 1-6.

Weir, B.S. 1996. Genetic data analysis 2: methods for discrete population genetic data. Sinauer, Massachusetts.

West, D.R., W.A. Compton, and M.A. Thomas. 1980. A comparison of replicated S1 per se vs. reciprocal full sib index selection in corn. I. Indirect response to population densities. Crop Sci. 20:35-42.

Wimmer, V., T. Albrecht, H.J., Auinger and C.C. Schoen. 2012. synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics. 28: 2086-2087.

2.8 Tables

Table 2.1. List of population abbreviations and corresponding descriptors.

Population	Base Population†	Selection Method ‡	Cycle	Replicate
NBS_C0	NBS	--	0	--
NBS_RFS_C8_1	NBS	RFS	8	1
NBS_RFS_C8_2	NBS	RFS	8	2
NBS_RFS_C8_3	NBS	RFS	8	3
NBS_S1_C8_1	NBS	S1	8	1
NBS_S1_C8_2	NBS	S1	8	2
NBS_S1_C8_3	NBS	S1	8	3
NSS_C0	NSS	--	0	--
NSS_RFS_C8_1	NSS	RFS	8	1
NSS_RFS_C8_2	NSS	RFS	8	2
NSS_RFS_C8_3	NSS	RFS	8	3
NSS_S1_C8_1	NSS	S1	8	1
NSS_S1_C8_2	NSS	S1	8	2
NSS_S1_C8_3	NSS	S1	8	3

†NBS = Nebraska B Synthetic; NSS = Nebraska Stiff Stalk Synthetic.

‡ RFS = Reciprocal full-sib selection; S1 = S1-progeny selection

Table 2.2. Number of polymorphic markers, observed heterozygosity (Ho), gene diversity (D) and realized effective population size (Ne) for the 14 UNL-RpRS populations. Polymorphic is defined as minor allele frequency ≥ 0.05.

Population	Polymorphic Markers	Ho	D	Ne
NBS_C0	382	0.322	0.310	--
NBS_RFS_C8_1	280	0.228	0.238	12
NBS_RFS_C8_2	250	0.200	0.191	9
NBS_RFS_C8_3	303	0.232	0.225	12
NBS_S1_C8_1	260	0.222	0.208	11
NBS_S1_C8_2	226	0.182	0.177	7
NBS_S1_C8_3	217	0.179	0.169	7
NSS_C0	459	0.356	0.340	--
NSS_RFS_C8_1	324	0.248	0.235	11
NSS_RFS_C8_2	293	0.259	0.246	13
NSS_RFS_C8_3	300	0.242	0.226	11
NSS_S1_C8_1	321	0.263	0.245	13
NSS_S1_C8_2	322	0.229	0.243	9
NSS_S1_C8_3	306	0.237	0.236	10
Std. Error	--	0.003	0.004	--

Table 2.3. Estimates of the Modified Rogers Distance (MRD) between pairs of populations from Nebraska B Synthetic (NBS) and Nebraska Stiff Stalk Synthetic (NSS) and standard errors. All markers were used to estimate MRD.MRD between complementary populations and populations undergoing S1-progeny selection in the same environments are underlined for ease of comparison.

	NSS						
Population	C0	RFS_C8_1	RFS_C8_2	RFS_C8_3	S1_C8_1	S1_C8_2	S1_C8_3
NBS_C0	0.34 ± 0.009	0.44 ± 0.011	0.38 ± 0.010	0.43 ± 0.011	0.42 ± 0.011	0.40 ± 0.010	0.39 ± 0.010
NBS_RFS_C8_1	0.38 ± 0.010	0.46 ± 0.012	0.42 ± 0.012	0.47 ± 0.012	0.44 ± 0.012	0.43 ± 0.011	0.42 ± 0.012
NBS_RFS_C8_2	0.42 ± 0.011	0.51 ± 0.013	0.45 ± 0.012	0.51 ± 0.013	0.48 ± 0.012	0.46 ± 0.012	0.46 ± 0.012
NBS_RFS_C8_3	0.40 ± 0.010	0.49 ± 0.012	0.44 ± 0.012	0.49 ± 0.012	0.46 ± 0.012	0.44 ± 0.012	0.44 ± 0.012
NBS_S1_C8_1	0.41 ± 0.010	0.50 ± 0.013	0.45 ± 0.012	0.50 ± 0.013	0.47 ± 0.013	0.46 ± 0.012	0.47 ± 0.012
NBS_S1_C8_2	0.43 ± 0.011	0.52 ± 0.013	0.47 ± 0.012	0.52 ± 0.013	0.49 ± 0.012	0.48 ± 0.012	0.47 ± 0.013
NBS_S1_C8_3	0.42 ± 0.011	0.51 ± 0.013	0.47 ± 0.013	0.51 ± 0.013	0.47 ± 0.013	0.48 ± 0.012	0.45 ± 0.013

Table 2.4. Estimates of the Modified Rogers Distance (MRD) between pairs of populations from Nebraska B Synthetic (NBS) and Nebraska Stiff Stalk Synthetic (NSS) and standard errors. Markers in linkage disequilibrium ($r^2 > 0.20$) with at least one adjacent locus were used to estimate MRD. MRD between complementary populations and populations undergoing S1-progeny selection in the same environments are underlined for ease of comparison.

	NSS						
Population	C0	RFS_C8_1	RFS_C8_2	RFS_C8_3	S1_C8_1	S1_C8_2	S1_C8_3
NBS_C0	0.37 ± 0.013	0.47 ± 0.016	0.39 ± 0.015	0.46 ± 0.016	0.43 ± 0.016	0.41 ± 0.014	0.38 ± 0.016
NBS_RFS_C8_1	0.40 ± 0.014	0.47 ± 0.017	0.43 ± 0.016	0.49 ± 0.016	0.46 ± 0.017	0.43 ± 0.016	0.42 ± 0.017
NBS_RFS_C8_2	0.46 ± 0.015	0.55 ± 0.019	0.47 ± 0.017	0.54 ± 0.018	0.51 ± 0.018	0.48 ± 0.018	0.47 ± 0.017
NBS_RFS_C8_3	0.42 ± 0.014	0.52 ± 0.019	0.44 ± 0.017	0.52 ± 0.017	0.47 ± 0.017	0.46 ± 0.017	0.43 ± 0.018
NBS_S1_C8_1	0.43 ± 0.016	0.53 ± 0.019	0.45 ± 0.018	0.51 ± 0.019	0.48 ± 0.019	0.45 ± 0.018	0.45 ± 0.018
NBS_S1_C8_2	0.46 ± 0.016	0.55 ± 0.018	0.48 ± 0.018	0.55 ± 0.018	0.53 ± 0.017	0.49 ± 0.017	0.47 ± 0.018
NBS_S1_C8_3	0.44 ± 0.016	0.52 ± 0.018	0.47 ± 0.019	0.54 ± 0.018	0.47 ± 0.019	0.50 ± 0.017	0.44 ± 0.018

2.9 Figures

Figure 2.1. Diagram illustrating structure of University of Nebraska Replicated Recurrent Selection program. The Nebraska Stiff Stalk (NSS) and Nebraska B Synthetic base populations were independently sampled six times, with three replicates subjected to S1-progeny selection (S_1) and Reciprocal Full-sib Selection (RFS).

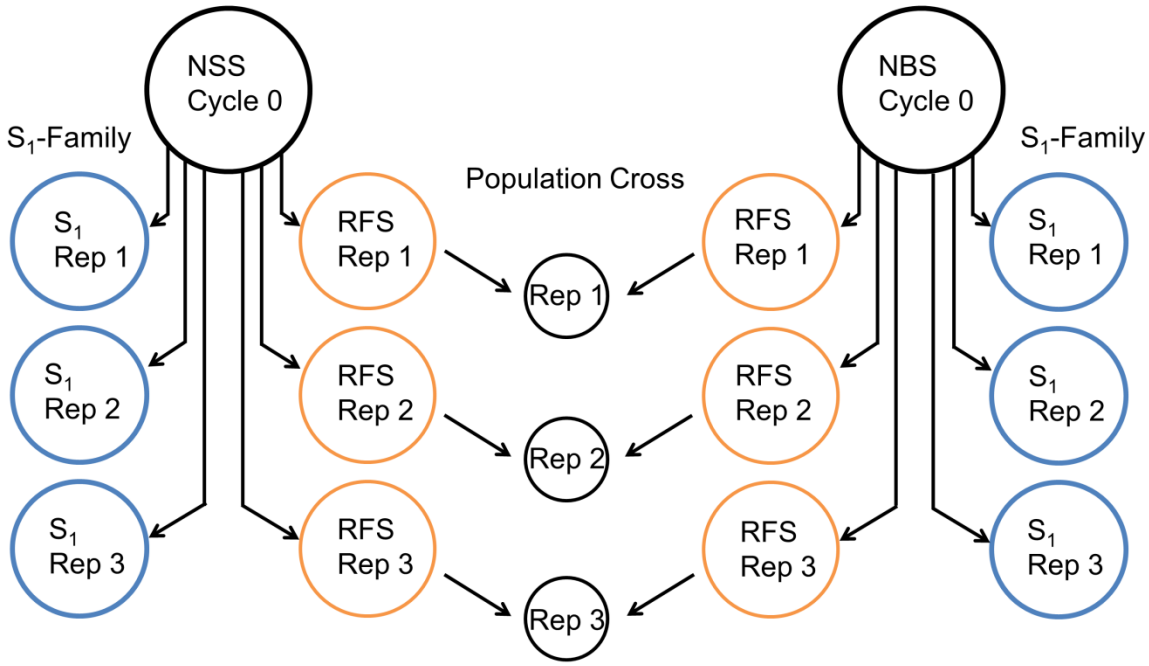


Figure 2.2. Column charts displaying the index value (units Mg/Ha) the populations comprising the University of Nebraska Replicated Recurrent Selection program as reported by three studies: Thomas (1979), Odhiambo (1987), and Galusha (1999). The number in parentheses next the study name indicates the highest cycle of selection evaluated by that study. The RFS panel displays population index values a population cross basis. The NSS-S1 and NBS-S1 panel display index values on a S1-progeny basis for the NSS and NBS populations, respectively. The numerical values displayed above sets of columns represent the index value averaged over the three replicate populations. An asterisk indicates that the displayed average value is significantly different than the Cycle 0 value ($P \leq 0.05$). A tilde above the RFS columns indicates the S1 and RFS populations were significantly different ($P \leq 0.05$) for those replicates.

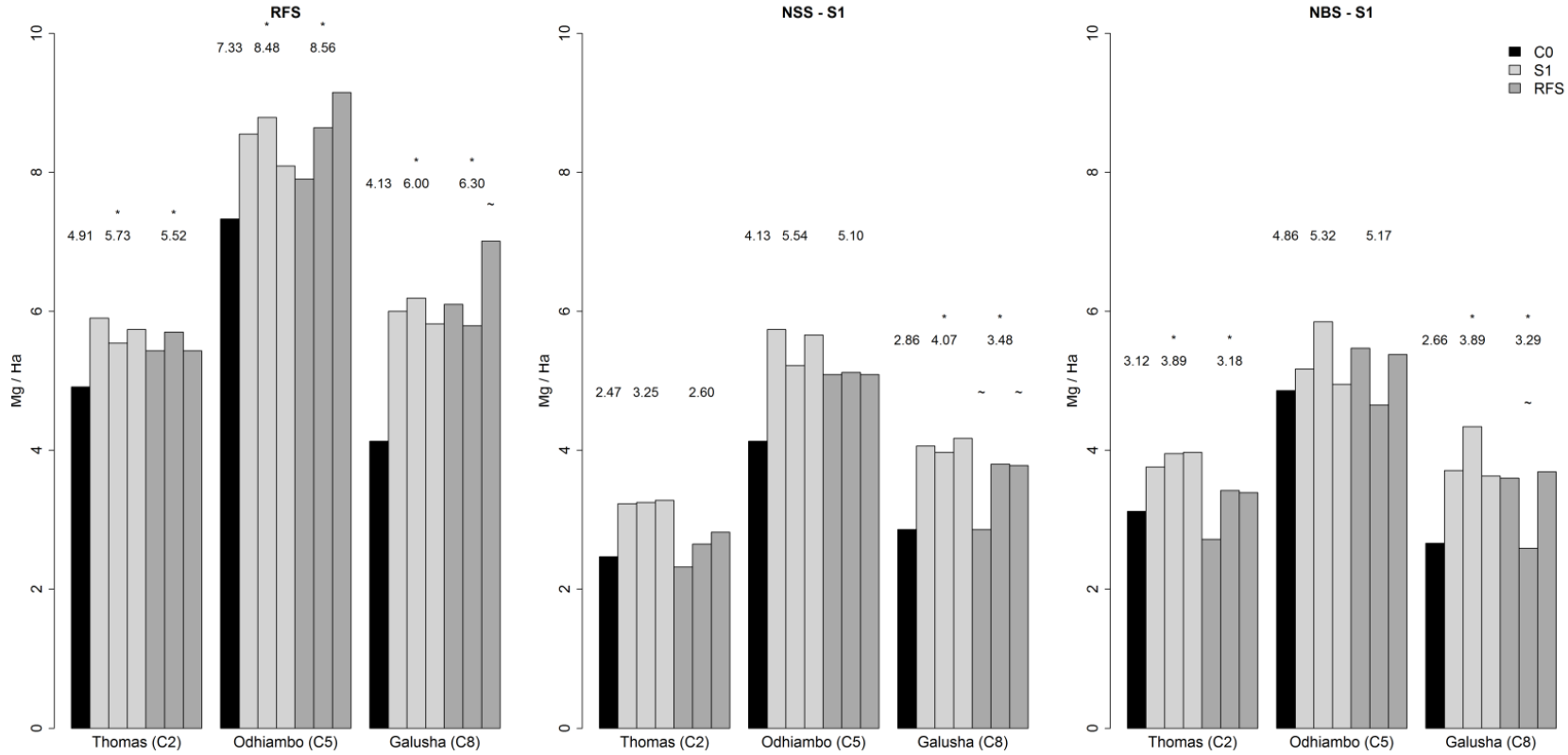

RFS
NSS - S1
NBS - S1
C0
S1
RFS
Mg / Ha
0
2
4
6
8
10
4.91 5.73 5.52
7.33 8.48 8.56
4.13 6.00 6.30
2.47 3.25 2.60
4.13 5.54 5.10
2.86 4.07 3.48
3.12 3.89 3.18
4.86 5.32 5.17
2.66 3.89 3.29
Thomas (C2)
Odhiambo (C5)
Galusha (C8)

Figure 2.3. F_{st} values between corresponding replicate populations (i.e., the same replicate number and selection method within a base population) are displayed down the center of the figure. F_{st} values among replicates within the same base population and selection method are displayed off to the side. Panel A displays F_{st} values estimated using all markers. Panel B displays F_{st} values estimated using markers in linkage disequilibrium ($r^2 > 0.20$) with at least one adjacent locus.

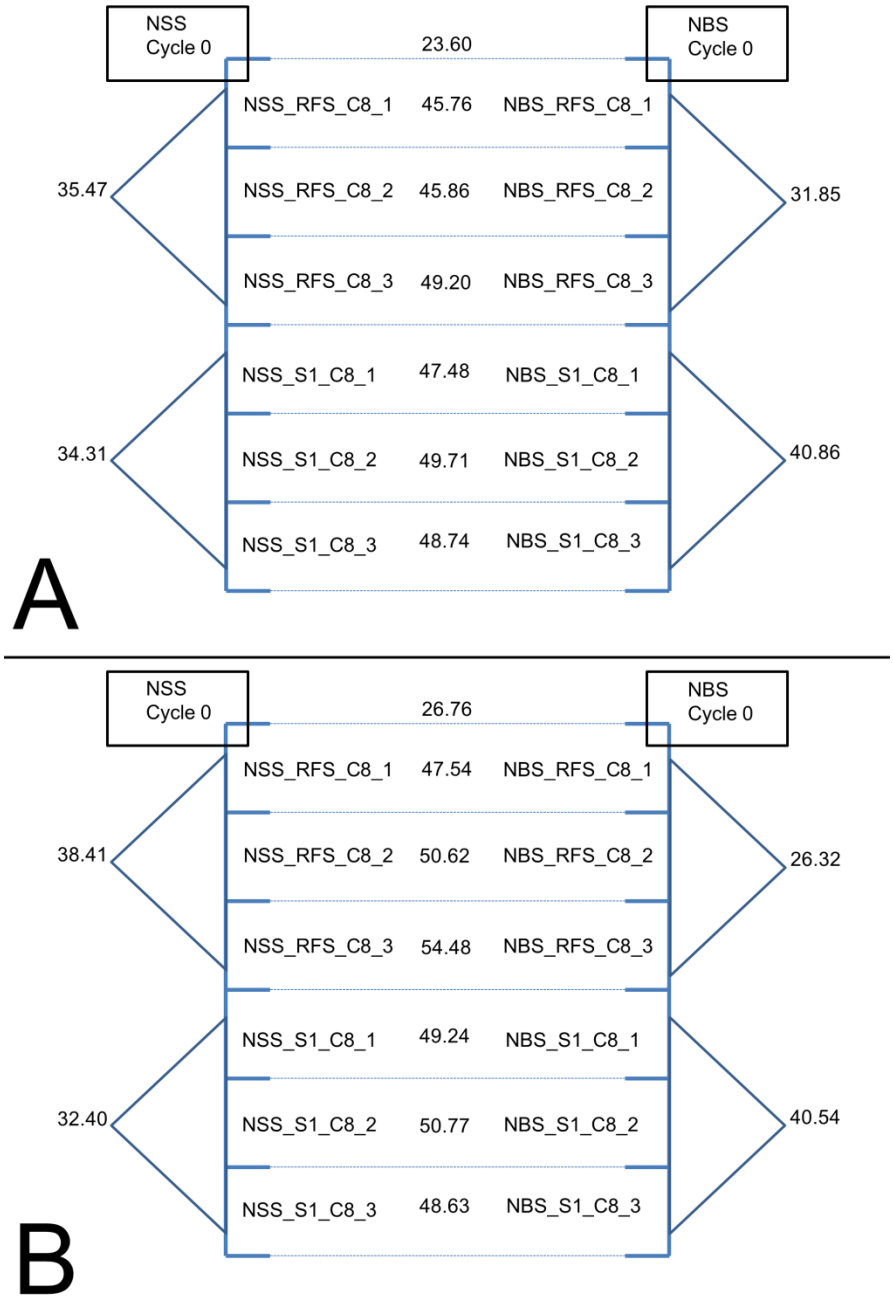
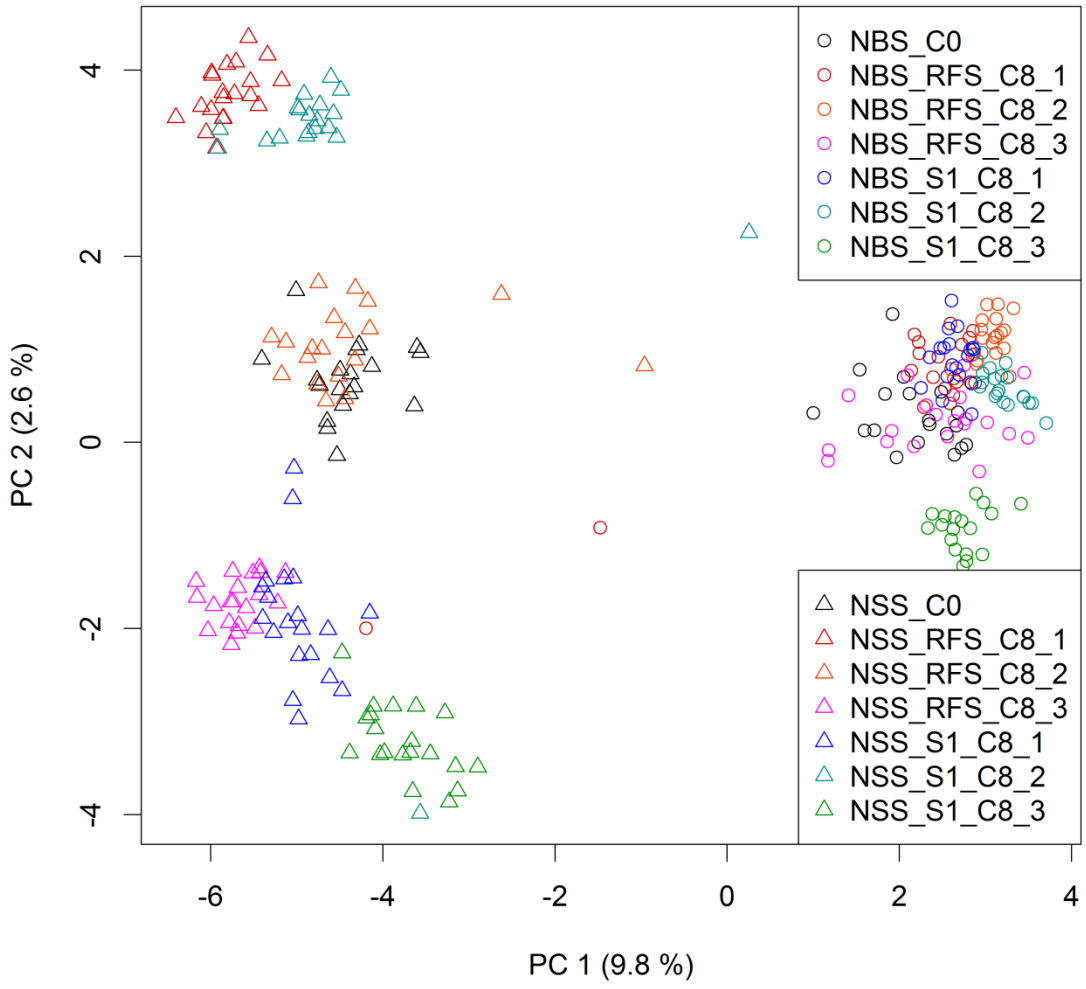


Figure 2.4. Principal component analysis for 14 populations in the University of Nebraska replicated recurrent selection program. Principal component 1 (PC 1) and PC 2 are plotted with the amount of variation explained by each PC in parentheses. Hot (red, orange, and pink) colors correspond to reciprocal full-sib selection, while cool (blue, teal, and green) colors correspond to S_1-progeny selection.



3 CHAPTER 3 PARALLEL EVOLUTION AND SELECTION FOR HYBRID PERFORMANCE IN THE UNIVERSITY OF NEBRASKA-LINCOLN REPLICATED RECURRENT SELECTION PROGRAM

3.1 Abstract

The inbred-hybrid system of maize breeding closely resembles a reciprocal full-sib (RFS) selection program. Studying changes in genetic variation as a result of RFS selection can help illuminate long-standing questions regarding the relative roles of selection and genetic drift and help understand the nature of adaptation occurring in selection programs. The University of Nebraska-Lincoln Replicated Recurrent Selection (UNL-RpRS) program underwent eight cycles of replicated RFS and S1-progeny selection, and is a powerful system to study the questions at hand. The objectives of this study were to identify regions of the genome under selection after eight cycles of selection and evaluate the effect eight cycles of selection for hybrid performance had in creating genome-wide and localized population structure. We address these questions with a large set of individuals sampled from the UNL-RpRS program with dense genotype-by-sequence data. We found evidence for parallel evolution in the UNL-RpRS program, with a region on chromosome 7 being implicated in hybrid performance and heterosis. Regions which evolve in parallel across independently run selection programs represent regions likely to be capitalizing on standing genetic variation and support a soft-sweep model of adaptation. No evidence for selection signatures was found which

could be attributed to selection methodology. This could be due to the nature of adaptation occurring in these populations or a result of unstable genetic topographies.

3.2 Introduction

How intense selection practiced in artificial selection programs shapes genetic variation is of immediate interest to plant breeders. And, more immediately relevant to corn breeders, is how genetic variation in reciprocal recurrent selection (RRS) programs shapes genetic variation. Reciprocal recurrent selection (Comstock et al., 1949) and its derivative reciprocal full-sib selection (RFS; Hallauer and Eberhart, 1970) were proposed as a method to increase performance of the population cross despite the mode of gene action underlying hybrid performance. The inbred-hybrid system of maize breeding resembles a RFS selection program (Duvick et al., 2004). Thus, RFS selection provides an excellent model system to study genetic changes resulting from selection for hybrid performance.

Several studies have examined genetic changes that occurred after RRS at the breeding program level (Labate et al., 1997; Hagdorn et al., 2003; Hinze et al., 2005; Romay et al., 2012; Gerke et al., 2013; Lamkey and Lorenz, 2014). In addition, other studies quantified genetic changes in large collections of historically important inbred lines and hybrids (Duvick et al., 2004; van Heerwaarden et al, 2012). Two trends are typically observed: 1) a reduction in genetic variation over time and, 2) the development of pronounced population structure. Naturally, these studies bring into question the

where, what and how regarding the effect selection has had in shaping genetic variation throughout the genome.

To this end, Stuber and Moll (1972) were among the first to use isozymes to track genetic changes after recurrent selection. More recently, Wisser et al. (2008) used single locus simulations inspired by Waples (1989) to identify selection targets for Northern Leaf Blight. Romay et al., (2012) evaluated a set of Spanish populations after three cycles of RFS selection and found four loci under selection. Beissinger et al., (2013) evaluated 30 cycles of mass selection for prolificacy and detected 28 regions under selection.

Genetic changes have also been examined extensively in the Iowa Stiff Stalk Synthetic (BSSS) and Iowa Corn Borer Synthetic (BSCB1) recurrent selection program (BSSS-BSCB1; Labate et al., 1997; Labate et al., 1999; Hinze et al., 2005; Gerke et al., 2013). Labate et al. (1999) identified 29 loci under selection in the BSSS and BSCB1 populations, one of which had increased the frequency of complementary alleles between the BSSS and BSCB1 populations. Labate et al. (1997) and Hinze et al. (2005) analyzed the genetic structure of the reciprocal recurrent selection program after 12 (Labate et al., 1997) and 15 (Hinze et al., 2005) cycles of selection and found genetic diversity decreased and was accompanied by an increase in population structure. More recently, Gerke et al. (2013) performed a genome-wide analysis of the BSSS-BSCB1 program using the Illumina MaizeSNP50 beadchip (Ganal et al. 2011) and detailed genome-wide simulations. These simulations show that with the exception of a few select regions, most of the population structure can be accounted for by genetic drift. The general consensus

from these studies was that a small fraction of the genome has been under detectable selection, rarely were target regions fixed in the genome, and genetic drift can be a very powerful force in defining population structure. The critical question is how powerful is selection relative to genetic drift?

Selection mapping offers an alternative approach to association analysis or traditional linkage mapping to identify regions of interest in the genome (Wisser et al., 2008) and may provide useful insights into the selection verse genetic drift question. Population genetics was a largely theoretical field prior to the advent of dense molecular marker data (Fisher, 1958; Crow and Kimura, 1970; Wright, 1978). Molecular marker data have granted scientists access to the genome simultaneously allowing decades of theoretical research to be formally tested (Weir and Cockerham, 1984; Waples, 1989; Jones et al., 2012). This field of research compares some proposed model of selection to a null hypothesis of random processes, namely genetic drift. If the selection signal is sufficiently strong the null hypothesis can be rejected and the candidate region is declared a selection candidate.

Methodology can be as simple (Stuber and Moll, 1972; Waples, 1989) or elaborate (Fariello et al., 2013) as an experiment allows. Selection mapping need not require population development outside the scope of a breeding program (Wisser et al., 2008; Yu et al., 2008), leverages population structure (Wright, 1978; Weir and Cockerham, 1984), and in some situations can identify signals associated with rare alleles. All three points can be problems with association mapping although methodology

has markedly improved (Zhao et al., 2007; McMullen et al., 2009; Manolio et al., 2009). Statistical power especially with the noise introduced by sequence data can be an issue, but can be overcome to a degree (Hohenlohe et al., 2010; Utsunomiya et al., 2013) or the two approaches can be considered together to increase power (Schwarzenbacher et al., 2012).

Here, we investigate genetic changes in the University of Nebraska-Lincoln Replicated Recurrent Selection program (UNL-RpRS) after eight cycles of RFS and S1-progeny selection for grain yield (Thomas, 1979; Odhiambo, 1987, Galusha, 1999). Lamkey and Lorenz (2014) analyzed genetic diversity after eight cycles of RFS and S1-progeny recurrent selection finding that i) as expected, genetic diversity decreased while population structure increased and, ii) using the Modified Roger's Distance (MRD) between S1-progeny selection programs as a null hypothesis for genetic drift, no difference in distance was observed between S1-progeny selection and RFS selection. Several questions remained that could not be addressed with the previous marker dataset (Jones et al., 2009). First, after eight cycles of selection which loci were under selection in the UNL-RpRS program? Second, what role did selection for hybrid performance have in diverging the NBS and NSS base populations? Here we wish to expand on those initial objectives by investigating these questions using dense SNP markers scored using genotype-by-sequencing (GBS). Specifically, the objectives of this study are to: 1) identify regions of the genome which have likely been under selection after eight cycles

of selection and, 2) evaluate the effect selection for hybrid performance has had in shaping population structure.

3.3 Materials and Methods

3.3.1 Genetic material

The UNL-RpRS program was previously summarized by Lamkey and Lorenz (2014). Briefly, RFS and S1-progeny selection were initiated from the Nebraska B Synthetic (NBS) and Nebraska Stiff Stalk Synthetic (NSS) populations with the goal of comparing two contrasting selection methods (Figure 2). On the one hand, S1-progeny selection is an intra-population improvement procedure (Fehr, 1991) designed to improve the performance of the population *per se* and capitalize on additive gene action. On the other hand, RFS selection is an inter-population improvement procedure (Fehr, 1991) designed to improve the population cross and expected to capitalize on additive genetic variation as well as dominant, overdominant, and pseudo-overdominant gene action (Comstock et al., 1949). Both selection methods, RFS and S1-progeny selection, were replicated three times from the base population. Same numbered replicates were used as the reciprocal population when making population crosses. Replicates were initiated in successive years to help manage the work load associated with running twelve simultaneous selection programs. Eight cycles of selection were conducted for a multiplicative index approximating machine harvestable yield (Galusha, 1999). The fourteen populations including sample size are summarized in Table 1, and a diagram illustrating the structure of the UNL-RpRS program is given in Supplementary Figure 1.

3.3.2 Sampling

Genetic sampling was conducted from balanced bulks developed from population increases grown in the Lincoln, NE summer nursery in 2013. Fourteen populations were sampled (Table 1). Ninety-five individuals were samples from the C0 populations and 47 individuals were sampled from the C8 populations.

3.3.3 DNA Extraction and Sequencing

The Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA) was used to extract DNA from C0 populations. The Qiagen BioSprint System (Qiagen, Valencia, CA) was used to extract DNA from all C8 populations. The switch in extraction platform was made due to the drastic time reduction afforded by the BioSprint platform. DNA samples were standardized to 100 ng ul^{-1} in 100 ul when possible. Otherwise, a minimum of 50 ng ul^{-1} in at least 50 ul was sought. Concentrations were determined on a Nanodrop 8000 (Thermo Scientific, Wilmington, DE). DNA quality was evaluated on 1 % agarose gels ran for 3.5 to 4 h by comparing to a DNA ladder digested with the *HindIII* restriction enzyme. Genotyping was performed at the Institute of Genomic Diversity (Cornell University, Ithaca, New York) using the GBS protocol outlined in Elshire et al. (2011). Genomic DNA was digested using the *ApeKI* restriction enzyme.

3.3.4 Bioinformatics

DNA read analysis and SNP calling was performed at University of Nebraska – Lincoln using TASSEL-GBS (Glaubitz, 2014). The Discovery Pipeline was implemented for this study, which was designed for use with a reference genome (Glaubitz, 2014). The

TASSEL-GBS pipeline takes for input raw sequence data, an indexed genome, and a barcode file and outputs SNP genotypes in either variant call format (VCF) or HapMap format. The hardmasked maize reference genome was downloaded from www.phytozome.net. The default settings were used for all plugins in the TASSEL-GBS pipeline with the exception of "MergeMultipleTagCounts" and "TagsToSNPByAlignment" (renamed DiscoveryCaller in Tassel 4) plugins. In the first case, a Tag (Glaubitz et al, 2014) had to appear at least 5 times across all sequence files to be included in the master tag file. In the second case, a minor allele frequency of 1 %, and the vcf flag was turned on to generate VCF files.

The VCF was chosen to hold the genotype data. First, because read depth and allele depth information is retained for each genotype and allele, respectively. Also, individuals in these populations are heterozygous, creating uncertainty in the genotype calls. Therefore, it makes intuitive sense to express confidence in the genotype calls via maximum likelihood. The genotyping method implemented with the vcf option in TASSEL-GBS follows Hohenlohe et al. (2010). An example of their approach is outlined in Hohenlohe et al. (2010).

A sample was considered failed if greater than 75 % of the data was missing. Ninety-five individuals were sampled from NBS_C0 and NSS_C0, but due to a high rate of missing data seven individuals were dropped from NBS_C0 and four individuals were dropped from NSS_C0. Forty-seven individuals were sampled from each of the 12 cycle

eight populations. Only two individuals were dropped due to high missing data from NSS_S1_C8_1 and NSS_S1_C8_2. In total, 741 individuals were retained for analysis.

VCF files were processed with VCFtools v0.12.1a (Danecek et al., 2011). To increase confidence in genotype calls, two different approaches were implemented. First, genotype depth, the total number of reads used to call a genotype, was filtered at values of 0, 2, 4, 5, 6, 8, and 10. Second, each genotype call in a VCF file is expressed as a likelihood function (Hohenloe et al., 2010) as discussed above. The likelihood can be expressed as a quality score which ranges from 0 to 100 where 100 represents perfect quality. Values of 25, 50, 75, 85, 95 and 98 were chosen. Each genotype depth or genotype quality value was assessed at 30, 40 or 50 % missing data (proportion of data missing) and a minor allele frequency of 1 %. Both C0 and C8 material were represented in the data and the C8 material is expected to display a higher inbreeding coefficient. For this step, each C0 population was analyzed separately. The final dataset was filtered with genotype quality of 75 and missing data of 40 %. This dataset had observed and expected heterozygosities which agreed well (Weir, 1996) and represented a good compromise between confidence in genotype calls, site depth, and marker number. The final dataset for analysis had 741 individuals and 312 102 SNPs.

3.3.5 Statistics

Both objectives of this study require estimates of allele frequency both within and across populations. Allele frequencies were estimated via maximum likelihood (Weir, 1996).

Wright's Fst (1978) was calculated with VCFtools v0.12.1a on a per locus basis using the methodology of Weir and Cockerham (1984), which accounts for unequal population sizes and sampling variance. Weir and Cockerham (1984) regard Fst as an estimator, not a population parameter. Since Fst is regarded as an estimator it can be undefined (i.e. 0/0) when a marker is monomorphic in all populations being compared, and Fst can also be negative with no biological interpretation. Undefined Fst estimates were set to zero for downstream processing and negative values were kept.

Modified Roger's Distance was used to quantify the genetic distance among populations and was computed as

$$MRD = \frac{1}{\sqrt{2m}} \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(p_{ij} - q_{ij}\right)^2}$$

where p_{ij} and q_{ij} are the allele frequencies of the *jth* allele at the *ith* locus, in each population, respectively. The standard error for each element of the MRD matrix can be obtained with a jackknife estimator by dropping a locus and computing the MRD (Efron and Tibshirani, 1986; Hagdorn, 2003). For each element of the distance matrix, the l^{th} locus was omitted and the MRD was computed. This was repeated *m* times for each element of the MRD matrix. The variance was found as:

$$V(MRD) = \frac{m-1}{m} \sum_{j=1}^{m} (MRD_{-j} - MRD_m)^2$$

where m represents the number of loci, $MRD_{\_j}$ represents the MRD with the j^{th} locus omitted, and MRD_m represents the mean of m estimates of $MRD_{\_j}$ (Efron and Tibshirani, 1986; Hagdorn et al., 2003).

To quantify genetic diversity, gene diversity was computed following Weir (1996) as,

$$D = 1 - \sum_{i=1}^{k} p_i^2$$

where k is the number of alleles and p_i^2 is the squared allele frequency. Genome-wide estimates were obtained by averaging across loci. A standard error of gene diversity was estimated by bootstrapping across loci 1,000 times.

To visualize genome-wide patterns of variation, eigenanalysis was conducted on the normalized genotype matrix following the methodology of Patterson et al. (2006). The first k significant eigenvalues were used to extract the first k eigenvectors. Eigenvectors two and higher did not have an obvious biological interpretation. The approach of Patterson et al. (2006) determines significance of eigenvalues by comparison to a Tracy-Widom distribution. The Tracy-Widom distribution was generated with the R package RMTstat (Johnstone et al., 2009).

3.3.6 Data Smoothing

A Gaussian kernel was fit to the Fst values computed in VCFtools following Hohenlohe et al. (2010). The kernel is computed as,

$$f(x) = exp\left(\frac{-(x-c)}{2\sigma^2}\right)$$

where x is the base pair position of the SNP in question, c is base pair position of the SNP in the center of the window and σ is the size of window in base pairs. Thus, Fst values within a window are weighted according to their proximity to the SNP in the center of the window, and as SNPs become distal to the center SNP, their weight decreases.

Windows of 500 Kb with a 125 Kb step were chosen by visual inspection. This approach was previously used by Hohenlohe et al. (2010) and is similar to other sliding window approaches based on number of SNPs per window used by other researchers (Jones et al., 2012; Beissinger et al., 2013). Since window size is based on the observed physical map, window size could vary around a mean window size of 500 Kb. Exceptional windows were defined as windows being more the two standard deviations away from the desired window size. This corresponded to windows greater than 722 082 bp and less than 277 918 bp.

3.3.7 Outlier test for identifying genomic regions putatively affected by selection

The distribution of windows from the kernel was compared to an outlier test at the 99.9 or 99.99 percentiles. Windows surpassing the 99.9 % and 99.99 % thresholds were considered candidates for strong and very strong selection, respectively. The outlier test is common in the population genetics literature (e.g. Akey et al., 2009; Beissinger et al., 2013).

Linkage disequilibrium (LD) in maize decays rapidly with a recent estimate of LD dropping to an r^2 value of 0.2 within 1 Kb based on a large genetically diverse collection of inbred lines (Romay et al., 2013). Within defined heterotic groups LD decay persisted to roughly 10 Kb (Romay et al., 2013). An LD based metric was not used here since the bulk of the information about LD structure in these populations is contained in the heterozygous genotype calls, creating the issue of phase between two loci in question when computing LD. Without information on the phase between two loci, LD is drastically underestimated (data not shown). Due to this issue, potential genes implicated in selection were identified as genes that fell within the 500 Kb window or partially overlapped the left or right side of the 500 Kb window.

3.4 Results

3.4.1 Genome coverage and sequencing depth

Prior to alignment, the TASSEL-GBS pipeline merges sequence reads into Tags, which serves as a quality control measure and also facilitates the alignment process (Glaubitz et al., 2014). A total of 9,148,273 Tags were aligned to the hardmasked B73 reference genome. Of these, 34.74 % aligned uniquely, 20.86 % percent aligned to multiple locations, and 44.40 % did not align. Based on uniquely aligned reads only, the TASSEL-GBS pipeline called over 1 000 000 SNPs among 741 individuals sampled across the 14 populations, resulting in over 741 000 000 data points. Filtering based on missing data, minor allele frequency, and genotype quality resulted in 312 102 SNPs. The 312 102 SNPs were well distributed across the genome with a median density of 60 SNPs per 500 Kb window (Figure 1A). The median depth per population ranged between 4x (NBS_C0) and 5x (NSS_S1_C8_2) and the median sequencing depth observed across the experiment was 3.9x with a range from 2x to 163x and a standard deviation of 2x (Table 2; Figure 1B). Individual depth was slightly higher with a mean of 4x and range of 4x (NBS_C0) to 5.5x (NBS_S1_C8_3; Table 2). Filtering with genotype quality greater than or equal to 75 resulted in an observed mean genotype quality of 84.6 (Figure 1C). Collectively, this information demonstrates that a dense physical map, with nearly 4x site coverage and good quality was built for genetic analysis.

3.4.2 Evidence of directional selection after eight cycles of RFS and S1 selection

A clear decrease in gene diversity and number of polymorphic markers can be seen between C0 and C8 populations (Table 2). Gene diversity was 0.132 and 0.142 in NBS_C0 and NSS_C0 base populations, respectively (Table 2). After eight cycles of RFS selection gene diversity in the NBS program dropped to a mean of 0.098, and gene diversity in the NSS program dropped to a mean of 0.109 (Table 2). Similarly, gene diversity in the S1-progeny selection program dropped to 0.100 and 0.119 for the NBS and NSS programs, respectively (Table 2). Evaluating gene diversity among replicate lines shows that diversity is preserved across replicates, but not completely. Among NBS_RFS and NBS_S1 replicates gene diversity was 0.116 and 0.119, respectively, and among NSS_RFS and NSS_S1 replicates gene diversity was 0.130 and 0.139, respectively. Since gene diversity was not completely preserved among replicate populations this means that similar regions of the genome were either fixed or lost during the course of the selection program. This relative loss of diversity is accompanied by an increase in genetic distance from C0 to C8 populations as evaluated by MRD (Table 3, bold).

To detect evidence for selection throughout the genome, a kernel smoothed Fst scan was implemented following Hohenlohe et al. (2010) and compared to an outlier test. Because the outlier test was used on datasets with the same number of windows, the same number of windows was identified from each dataset. Sixty-four windows were identified

with the outlier test across the RFS and S1 selection programs at the 99.9 % level and 8 remained at the 99.99 % level.

To focus the results of this section, the intersections between replicate populations from the same base population and selection method were examined. These regions have been highlighted in orange in Figures 2 to 5. The windows contained in these intersections represent regions that have diverged in the same manner across independent selection programs for machine-harvestable yield. Generally, a small (n <= 3) number of windows were shared between any two replicates and no program shared a common set of loci among the three replicates. A complete list of selection candidates can be found in Supplementary Table 1.

The NBS_RFS set of populations shared no windows in common (Figure 2). The replicates of the NBS_S1 selection program identified a window which was on chromosome 1:176170056-176757342 and implicated gene GRMZM2G095469, which is not characterized (Figure 3). Three windows overlapped among the replicates of the NSS_RFS selection program. Three windows overlapped between the first and second replicates identifying windows on chromosomes 2:86916123-87656448, 3:73784151-74096740, and 6:30257596-30516029 (Figure 4). One window overlapped between the second and third replicates identifying a region on chromosome 7:90246776-90789748 (Figure 4). The region on chromosome 2 implicated three gene models GRMZM2G127457, GRMZM2G028730, and GRMZM2G154558. Only GRMZM2G154558 was associated with tranferase activity in the genome. The region on

chromosome 3 implicated GRMZM2G701566, which is an expressed protein with no known function. Finally, the regions on chromosomes six and seven did not overlap with any gene models. The NSS_S1 selection program identified a shared window between the second and third replicates on chromosome 6:101192553-101513634 but did not overlap with a gene model (Figure 5).

Of particular interest were two adjacent regions on chromosome 7 (113535012:114178113 bp and 113706281:114196433 bp) which have been previously associated with selection for hybrid performance (Labate et al., 1999) and heterosis (Schon et al., 2010). This region was found to be under selection in both S1 programs (Figures 3 and 5), but only in the NBS_RFS_C8_3 program (Figure 2C). The region identified on chromosome 7 does not overlap any gene models from the B73 genome.

Pooling information among replicate cycle eight populations can help emphasize which regions show a clear and consistent change in allele frequency between cycle zero and cycle eight. Of the 10 regions previously discussed, four also passed the outlier test in the pooled comparisons. The two adjacent regions on chromosome 7 (113535012:114178113 bp and 113706281:114196433 bp) appeared in, NBS_S1 (Figure 3D), NSS_RFS (Figure 4D), and NSS_S1 (Figure 5D) comparisons. Additionally, the region on chromosome 2 (86916123:87656448 bp) appeared in the NS_RFS comparison (Figure 5D).

3.4.3 Effect of selection on hybrid performance in diverging base populations

The second objective of this study was to explore the effect selection for hybrid performance had, if any, on the formation of population structure. If selection for hybrid performance has been effective at fixing complementary alleles in reciprocally selected populations due to overdominance or pseudooverdominant gene action, the distance between reciprocally selected RFS populations is expected to be greater than that observed between S1 populations (Lamkey and Lorenz, 2014). Since the selection target of S1-progeny selection is the performance of the population *per se*, the increased distance between reciprocal S1-progeny populations is likely due to genetic drift.

In examining the MRD between reciprocally selected populations, no apparent differences were observed between the RFS and S1 comparisons for same numbered replicates (Table 3, bold). Concomitantly, gene diversity dropped 35 % and 30 % in NBS_RFS and NSS_RFS populations, respectively; and dropped 32 % and 19 % in the NBS_S1 and NSS_S1 populations, respectively. These results agree well with previous studies investigating the effects of RRS on population structure and genetic diversity (Hinze et al. 2005; Lamkey and Lorenz, 2014; Gerke et al., 2013). If RFS selection had been effective in diverging a large proportion of the genome more than expected by chance, more pronounced population structure would be observed between reciprocally selected populations in the eigenvector plot (Figure 6). The first eigenvector explains variation (Eigen value = 5 %) differentiates the NSS and NBS base populations. No obvious trend is apparent for the second Eigen vector (Eigen value = 2 %). Most

famously the BSSS-BSCB1 RRS program observed very pronounced population structure in various evaluations of the selection program (Hinze et al., 2005; Gerke et al., 2013). However, approximately twice the number of cycles of selection had been conducted (Gerke et al., 2013), and sets of intervening cycles is available.

To examine the effect RFS selection had throughout the genome, a kernel smoothed Fst value was used as before. Since NBS_C0 and NSS_C0 were already highly diverged (MRD = 0.116; Table 3) prior to selection, the absolute difference in Fst between the cycle zero populations was compared to the absolute difference in Fst between C8 populations. This approach identified 32 regions at the 99.9 % level and 8 remained at the 99.99 % level for each selection method. The selection candidates from RFS selection were associated with 54 genes at the 99.9 % level and S1 selection was associated with 88 genes at the 99.9 % level.

Applying the same reasoning as presented in Lamkey and Lorenz (2014), trends were searched for across the genome that might point to differences between the selection methods. Results from these selection scans were very similar (Figures 7 and 8) suggesting that across the genome the effect for selection for inter-population performance did not leave a pronounced selection signature on the genome. If selection for hybrid performance had been effective, it would be reasonable to think that a build up of repulsion phase linkages in or near the centromeric regions would be expected (Hill and Robertson, 1966; McMullen et al., 2009). To test this, the absolute value of the distance of each window passing the outlier test at 99.9 % was measured from the

centromere on the appropriate chromosome. The results were averaged over chromosomes for each selection method and compared with a t-test. No significant differences in location were identified between the selection methods (data not shown).

3.5 Discussion

This study, coupled with previous research (Lamkey and Lorenz, 2014), shows that after eight cycles of selection gene diversity decreased along with the number of polymorphic markers, while genetic differentiation between populations increased after eight cycles of selection. These results are consistent with studies examining the effect of RRS methodologies on the genetic variation of maize populations (Labate et al. 1997; Hinze et al., 2005; Gerke et al., 2013; Lamkey and Lorenz, 2014) and also with trends observed across time points in large samples of maize inbred lines (Duvick et al., 2004; van Heerwaarden et al., 2012). Of interest is the lack of structure due to selection in the Eigen vector plot (Figure 6). The first Eigen vector explains variation due to base populations, which were structured prior to the start of selection. However, the second Eigen vector has no obvious biological interpretation. This is in contrast to the very explicit population structure observed in the BSSS-BSCB1 RRS program (Gerke et al., 2013). One important caveat to this result is that the BSSS-BSCB1 RRS program always had multiple cycles of selection, which can help delineate genetic structure. To help interpret these results it could be useful to review some introductory quantitative genetic theory.

Forces which act to change allele frequencies can be summarized as systematic and dispersive (Falconer and Mackay, 1996). Systematic forces (e.g. selection, mutation and migration) act to change allele frequencies in a manner predictable in amount and direction, while dispersive forces (e.g. genetic drift) act to change allele frequencies in a manner only predictable in magnitude. Population genetics was developed assuming the absence of all forces (Stephens, 2004). To understand the effect each force has on the population, each force is considered in turn by relaxing one or more assumptions. Eventually leading to the development of elaborate theory to help determine how multiple forces acting together change allele frequencies in populations (Crow and Kimura, 1970). In recurrent selection programs, the time frame is generally short enough to discard the effects of mutation, although exceptions do exist (Dudley and Lambert, 1992) and since mating is tightly monitored, migration is not a concern. This leaves selection and genetic drift as the two forces responsible for allele frequency change.

Drift has been demonstrated to play a very substantial role in the evolution of artificial selection systems. Both simulation based studies (Gerke et al., 2013) and empirical studies (Lamkey and Lorenz, 2014) demonstrate that essentially all the observed population structure in maize recurrent selection programs is due to genetic drift. This observation is corroborated by a generation means analysis performed after five cycles of selection in the UNL-RpRS program (Tragesser et al., 1989). Tragesser et al. (1989) found genetic drift and selection to be significant sources of variation and equal, but opposite in direction.

An advantage of the UNL-RpRS program is the built in replication of the selection methodology. This can provide clues to the nature of genetic drift and lend support to regions under selection. Previous research (Lamkey and Lorenz, 2014) measured variation among sets of replicate populations and found the degree of divergence consistent with what is expected under single locus theory. In the absence of other forces, single locus theory predicts that in a large but finite population, which becomes subdivided will eventually redistribute genetic variation from within populations to among populations (Falconer and Mackay, 1996). Ultimately, no genetic variation is lost; rather it is redistributed across subpopulations. Gene diversity measured in the UNL-RpRS program computed across replicate populations restored much of the genetic diversity which was lost within replicates. The remaining fraction of genetic diversity that was not restored could represent regions where selection has fixed similar alleles across the replicate populations, or due to three replicates not adequately capturing variation among populations. In several instances similar regions were increased in frequency between the replicates (Figures 2-5).

The role of selection is more complicated. Hard sweep (Maynard Smith and Haigh, 1974) and soft sweep (Hermisson and Pennings, 2005) models are used to describe how populations adapt to new selection pressures. In maize, selection response is usually observed immediately (West et al., 1980; Coors, 1999) and recent research in maize using dense genome-wide marker data in a recurrent selection program has demonstrated selection is consistent with a soft sweep model in several instances

(Beissinger et al., 2013). Soft sweeps present a difficult and interesting challenge in the search for selection candidates (Sabeti et al., 2002; Innan and Kim, 2004; Hermisson and Pennings, 2005; Przeworski et al., 2005; Stephens and Scheet, 2005; Barrett and Schluter, 2008). Soft sweeps represent a model of adaptation which can occur from standing genetic variation in populations. They imply that prior to selection, recombination has ample opportunity to distribute the potentially favorable alleles throughout the population.

A search to find putative candidates for directional selection revealed a total of 128 regions from the RFS and S1 selection methods (Supplementary Table 1). Of the 128 selection candidates, 10 regions overlapped between two different replicates from each selection method – base population combination (NSS_RFS, NSS_S1, NBS_S1) with the exception of the NBS_RFS populations. Of particular interest are two adjacent regions on chromosome 7 between 113535012:114178113 bp and 113706281:114196433 bp. This region of chromosome 7 has been previously associated with selection for hybrid performance (Labate et al., 1999) and heterosis (Schon et al., 2010). Interestingly, no gene models overlapped with the windows identified in our study. The marker identified by Labate et al. (1999), *bnl15.40*, does not have physical coordinates, but searching www.maizegdb.org for flanking markers creates an ~ 5 Mb window containing over 20 gene models. Selection on non-genic DNA has been shown to modify phenotype (Clark et al., 2006), and selection on non-genic regions has been documented (Beissinger et al., 2013). Further research is required to understand the region under selection, in terms of

functional role in plant development and distribution of causal variants in the UNL-RpRS populations.

The remaining regions are more difficult to interpret. Duvick et al. (2004) suggested that selection in the inbred-hybrid system of maize breeding for hybrid performance selects phenotypes which are tolerant to biotic and abiotic stresses. This mode of adaptation increases or maintains yield in adverse environments. Recent work (Edwards, 2011) in the BSSS recurrent selection program on plant morphology revealed that phenotype has changed in parallel with cycles of selection. After 17 cycles of selection in BSSS tassel branch number decreased and leaf architecture changed from a broad, droopy orientation to a slight, vertical orientation. Although, the remaining regions identified in this study contained gene models the function was either undocumented or ambiguous with respect to the selection parameters of the UNL-RpRS program. As genotyping and annotation strategies improve in the maize genome it will become possible to better localize and understand how these regions contributed to increased yield.

From a practical standpoint, the common regions identified here remain ambiguous. From the population genetics perspective, the 10 regions identified here represent instances of parallel evolution. That is, regions which have independently come under selection in independent selection programs, and provide clues to the underlying mechanisms of adaptation. The identification of parallel targets of selection is not new. Extensive research has been performed in the three spine stickleback (*Gasterosteus*

aculeatus), a small oceanic fish which has an armored phenotype, a heavy bone structure on its underside, in marine environments (Barret et al., 2008). Following the most recent round of glacial retreat the three spine stickleback colonized freshwater environments across the Northern Hemisphere, and subsequently lost the armored phenotype to varying degrees (Barret et al., 2008; see Figure 1 of that paper for an example of the different morphs). Research has been conducted to characterize these regions (Hohenlohe et al., 2010; Jones et al., 2012) with the general conclusion that adaptation is consistent with a soft sweep model of evolution because the rapid and ubiquitous change in these regions is observed across freshwater populations. Furthermore, the major allele in freshwater populations is present in marine populations at a minor allele frequency (Barret et al., 2008; Hohenlohe et al., 2010).

Identifying parallel regions across replicate populations and in the pooled cycle eight population lend support to the hypothesis that parallel change is occurring in the UNL-RpRS because consistent allele frequency differences must be observed across all three replicate populations to observe a strong selection signal. Considering the forces of selection and genetic drift simultaneously introduces the concept of fixation probabilities (Crow and Kimura, 1970; Hermisson and Pennings, 2005). Fixation probabilities were originally derived considering a single new mutation in the population (Crow and Kimura, 1970). For an allele to become fixed, its fate depends on the strength of selection to fix the allele versus the power genetic drift has to remove the allele from the population (Walsh, 2004), which can be significant for new mutations. Considering soft

sweeps, the situation is more complicated because the genetic variation was present prior to selection. These regions could have also previously been under selection prior to the new selection regime, or could be neutral or even slightly deleterious prior to selection (Orr and Betancourt, 2003; Hermission and Pennings, 2005). Regardless, these regions represent good candidates for soft sweeps and present evidence of parallel adaptation in these recurrent selection programs. Research is underway to better characterize the identified parallel regions in the UNL-RpRS program with more sophisticated statistical machinery following Jones et al. (2012).

The second part of this study attempted to address the effect selection for hybrid performance has had in diverging base populations. Here, across both selection methods 32 regions were flagged as selection candidates. Examining the distribution of selection candidates between the methods does not reveal a distinct pattern of variation associated with selection method. Several studies have been conducted in maize, which have found evidence for selection (Labate et al., 1999; Romay et al., 2012; van Heerwaarden et al., 2012; Beissinger et al., 2013), but recent studies found the observed population structure can be attributed to genetic drift (Gerke et al., 2013; Lamkey and Lorenz, 2014). Multiple lines of reasoning can be put forth to explain the similar selection patterns observed between the RFS and S1-progeny selection programs. One explanation exists which can be considered an extreme case of the soft sweep model of adaptation termed polygenetic adaptation (Pritchard et al., 2010). Polygenetic adaptation implies allele frequency change to such a small degree that Δp (Falconer and Mackay, 1996) is undetectable by

most conventional selection mapping machinery particularly with small selection coefficients. While allele frequency changes are small, collectively, the change is pronounced enough to yield a sufficiently new phenotype in the population. Polygenetic adaptation provides a powerful robust mechanism for adaptation with few loci becoming differentiated to a high enough degree to be detected between the selection methods.

A second explanation starts with the basic assumptions Comstock et al. (1949) made about RRS. The authors derived their results for two highly diverged base populations assuming a two allele genetic system. Most likely, in two highly diverged populations prior to the start of selection, more than two alleles will exists across populations for any given locus. Research done by Griffing (1963) suggests that when two populations are crossed to evaluate hybrid performance, with more than two alleles, an unstable genetic "topography" is observed in the hybrid. In other words, at meiosis each of the alleles present has an equal and random chance of being sampled to construct the genotype of the hybrid. If different alleles are sampled at each round of selection for hybrid performance, the power of RRS to fix complementary alleles in populations is effectively negated. This is because with more than two alleles, the genotype of the hybrid can never be representative of the gametic array produced by populations with more than two alleles at a locus.

Cress (1967) suggested that, prior to the start of selection all desired germplasm should be intermated to form a large synthetic population. To initiate selection, the newly formed synthetic is randomly sampled to create A and B populations. One critical, and

expected, role of genetic drift could initially reduce variability within the sampled populations and create variance between populations at the onset of the RRS program. After a sufficient amount of time and divergence, RRS can act to increase the frequency of alleles which have risen to intermediate frequency in the reciprocal populations (Cress, 1967; Tracy and Chandler, 2004). Thus, even though significant selection response is ubiquitously observed across numerous selection programs (Coors, 1999), the ultimate potential might be initially defined by chance. Previous analysis of the UNL-RpRS program found that after eight cycles of selection, the third replicate of the RFS selection program outperformed the other replicates in the population cross (Galusha, 1999; Lamkey and Lorenz, 2014). Butruille et al. (2004) put Cress's (1967) idea into practice in the Wisconsin Golden Glow population. Reciprocal full-sib selection was practiced exactly in the way Cress (1967) suggested. After six cycles of selection, they found a significant increase in the population cross yield, and an increase in genetic differentiation, but were unable to attribute the difference to selection or genetic drift (Butruille et al., 2004).

3.6 Conclusions

We conducted a selection mapping study which identified 10 regions of the maize genome which have undergone parallel adaptation for an index approximating machine harvestable yield. These regions contain alleles that have increased in frequency after eight cycles of selection from standing genetic variation. One region on chromosome 7 was implicated with selection for hybrid performance, and represents a promising region

for future study. We did not detect genome-wide or spatial patterns of genetic variation unique to S1-progeny or RFS selection. This could be due to unstable genetic topographies in the population cross or simply due to the nature of adaptation at work in maize recurrent selection programs.

3.7 References

Akey, J.M. 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19:711-722.

Barrett, R.D., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends in Ecology & Evolution 23:38-44.

Barrett, R.D., S.M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. Science 322:255-257.

Beissinger, T.M., C.N. Hirsch, B. Vaillancourt, S. Deshpande, K. Barry, C.R. Buell, S.M. Kaeppler, D. Gianola, and N. de Leon. 2014. A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. Genetics 196:829-840.

Butruille, D.V., H.D. Silva, S.M. Kaeppler, and J.G. Coors. 2004. Response to selection and genetic drift in three populations derived from the golden glow maize population. Crop Sci. 44:1527-1534.

Clark, R.M., T.N. Wagler, P. Quijada, and J. Doebley. 2006. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. Nat. Genet. 38:594-597.

Comstock, R.E., H. Robinson, and P. Harvey. 1949. A breeding procedure designed to make maximum use of both general and specific combining ability. Agron. J. 41:360-367.

Coors, J.G. 1999. Selection methodology and heterosis. The Genetics and Exploitation of Heterosis in Crops 225-245.

Cress, C. 1967. Reciprocal recurrent selection and modifications in simulated populations. Crop Sci. 7:561-567.

Crow, J.F., and M. Kimura. 1970. An introduction to population genetics theory. The Blackburn Press, Caldwell, New Jersey.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Dudley, J., and R. Lambert. 1992. Ninety generations of selection for oil and protein in maize. Maydica 37:1-7.

Duvick, D.N., J. Smith, and M. Cooper. 2004. Long-term selection in a commercial hybrid maize breeding program. Plant Breed. Rev. 24:109-152.

Edwards, J. 2011. Changes in plant morphology in response to recurrent selection in the iowa stiff stalk synthetic maize population. Crop Sci. 51:2352-2361.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science. 1:54-77.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One 6:e19379.

Falconer, D., and T. Mackay. 1996. Introduction to Quantitative Genetics. Pearson Prentice Hall, Essex, England.

Fariello, M.I., S. Boitard, H. Naya, M. SanCristobal, and B. Servin. 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193:929-941.

Fehr, W.R. 1991. Principles of Cultivar Development. Iowa State University, Ames, IA.

Fisher, R.A. 1958. The genetic theory of natural selection. Dover.

Galusha, D.D. 1999. An evaluation of eight cycles of replicated recurrent selection in maize. Ph.D. diss. Univ. of Nebr., Lincoln, NE.

Ganal, M.W., G. Durstewitz, A. Polley, A. Bérard, E.S. Buckler, A. Charcosset, J.D. Clarke, E. Graner, M. Hansen, and J. Joets. 2011. A large maize (zea mays L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PloS One 6:e28334.

Gerke, J.P., J.W. Edwards, K.E. Guill, J. Ross-Ibarra, and M.D. McMullen. 2013. The genomic impacts of drift and selection for hybrid performance in maize. ArXiv Preprint arXiv:1307.7313.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PloS One 9:e90346.

Griffing, B. 1963. Comparisons of potentials for general combining ability selection methods utilizing one or two random-mating populations. Aust. J. Biol. Sci. 16:838-862.

Hagdorn, S., K.R. Lamkey, M. Frisch, P.E. Guimaraes, and A.E. Melchinger. 2003. Molecular genetic diversity among progenitors and derived elite lines of BSSS and BSCB1 maize populations. Crop Sci. 43:474-482.

Hallauer, A.R., and S. Eberhart. 1970. Reciprocal full-sib selection. Crop Sci. 10:315-316.

Hermisson, J., and P.S. Pennings. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335-2352.

Hill, W., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269-294.

Hinze, L.L., S. Kresovich, J.D. Nason, and K.R. Lamkey. 2005. Population genetic diversity in a maize reciprocal recurrent selection program. Crop Sci. 45:2435-2442.

Hohenlohe, P.A., S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson, and W.A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics 6:e1000862.

Innan, H., and Y. Kim. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl. Acad. Sci. U. S. A. 101:10667-10672.

Johnstone, I., Z. Ma, P. Perry, and M. Shahram. 2009. RMTstat: Distributions, statistics and tests derived from random matrix theory. R Package Version 0.2.

Jones, E., W. Chu, M. Ayele, J. Ho, E. Bruggeman, K. Yourstone, A. Rafalski, O.S. Smith, M.D. McMullen, and C. Bezawada. 2009. Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (zea mays L.) germplasm. Mol. Breed. 24:165-176.

Jones, F.C., M.G. Grabherr, Y.F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M.C. Zody, and S. White. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55-61.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1999. Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor. Appl. Genet. 99:1166-1178.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1997. Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. 37:416-423.

Lamkey, C., and A. Lorenz. 2014. Relative effect of drift and selection in diverging populations within a reciprocal recurrent selection program. Crop Sci. 54:576-585.

Manolio, T.A., F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, and A. Chakravarti. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753.

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, P. Brown, C. Browne, M. Eller, K. Guill, C. Harjes, D. Kroon, N. Lepak, S.E. Mitchell, B. Peterson, G. Pressoir, S. Romero, M. Oropeza Rosas, S. Salvo, H. Yates, M. Hanson, E. Jones, S. Smith, J.C. Glaubitz, M.

Goodman, D. Ware, J.B. Holland, and E.S. Buckler. 2009. Genetic properties of the maize nested association mapping population. Science 325:737-740.

Odhiambo, M.O. 1987. A comparison of replicated S1 line per se and reciprocal full-sib index selection in maize (zea mays L.). Ph.D. diss. Univ. of Nebr., Lincoln, NE.

Orr, H.A., and A.J. Betancourt. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics 157:875-884.

Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genetics 2:e190.

Prezeworski, M., G. Coop, and J.D. Wall. 2005. The signature of positive selection on standing genetic variation. Evolution 59:2312-2323.

Pritchard, J.K., J.K. Pickrell, and G. Coop. 2010. The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. Current Biology 20:R208-R215.

Romay, M.C., A. Butrón, A. Ordás, P. Revilla, and B. Ordás. 2012. Effect of recurrent selection on the genetic structure of two broad-based spanish maize populations. Crop Sci. 52:1493-1502.

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, R.J. Elshire, C.B. Acharya, S.E. Mitchell, and S.A. Flint-Garcia. 2013.

Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14:R55.

Sabeti, P.C., D.E. Reich, J.M. Higgins, H.Z. Levine, D.J. Richter, S.F. Schaffner, S.B. Gabriel, J.V. Platko, N.J. Patterson, and G.J. McDonald. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.

Schön, C.C., B.S. Dhillon, H.F. Utz, and A.E. Melchinger. 2010. High congruency of QTL positions for heterosis of grain yield in three crosses of maize. Theor. Appl. Genet. 120:321-332.

Schwarzenbacher, H., M. Dolezal, K. Flisikowski, F. Seefried, C. Wurmser, C. Schlotterer, and R. Fries. 2012. Combining evidence of selection with association analysis increases power to detect regions influencing complex traits in dairy cattle. BMC Genomics 13:48-2164-13-48.

Smith, J.M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23-35.

Stephens, C. 2004. Selection, drift, and the “Forces” of evolution. Philosophy of Science 71:550-570.

Stephens, M., and P. Scheet. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. The American Journal of Human Genetics 76:449-462.

Stuber, C., and R. Moll. 1972. Frequency changes of isozyme alleles in a selection experiment for grain yield in maize (zea mays L.). Crop Sci. 12:337-340.

Thomas, M.A. 1979. Replicated S1 line per se and reciprocal full-sib selection compared in maize (zea mays L.). Ph.D. diss. Univ. of Nebr., Lincoln, NE.

Tracy, W., and M. Chandler. 2006. The historical and biological basis of the concept of heterotic patterns in corn belt dent maize. p. 219-233. *In* K.R. Lamkey and M. Lee (ed.) Plant breeding: The Arnel R. Hallauer International Symposium. Blackwell Publishing, Ames, IA.

Tragesser, S., W. Youngquist, O. Smith, and W. Compton. 1989. Drift vs selection effects from 5 recurrent selection programs in maize. Maydica 34:23-32.

Utsunomiya, Y.T., A.M.P. O'Brien, T.S. Sonstegard, C.P. Van Tassell, A.S. do Carmo, G. Meszaros, J. Soelkner, and J.F. Garcia. 2013. Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. PloS One 8:e64280.

van Heerwaarden, J., M.B. Hufford, and J. Ross-Ibarra. 2012. Historical genomics of north american maize. Proc. Natl. Acad. Sci. 109:12420-12425.

Walsh, B. 2004. Population-and quantitative-genetic models of selection limits. Plant Breeding Reviews. 24.1:177-225.

Waples, R.S. 1989. Temporal variation in allele frequencies: Testing the right hypothesis. Evolution 1236-1251.

Weir, B.S., and C. Cockerham. 1996. Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc. Inc., Sunderland, MA, USA.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 1358-1370.

West, D., W. Compton, and M. Thomas. 1980. A comparison of replicated S1 per se vs. reciprocal full-sib index selection in corn. I. indirect response to population densities. Crop Sci. 20:35-42.

Wisser, R.J., S.C. Murray, J.M. Kolkman, H. Ceballos, and R.J. Nelson. 2008. Selection mapping of loci for quantitative disease resistance in a diverse maize population. Genetics 180:583-599.

Wright, S. 1978. Vol. 4: Variability within and among natural populations. University of Chicago Press, Chicago, IL.

Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178:539-551.

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, and P. Marjoram. 2007. An arabidopsis example of association mapping in structured samples. PLoS Genetics 3:e4.

3.8 Tables

Table 3.1. List of population abbreviations, corresponding descriptors, and sample size per population.

Population	Base pop.	Method	Cycle	Replicate	Sample size
NBS_C0	NBS	-	0	-	88
NBS_RFS_C8_1	NBS	RFS	8	1	47
NBS_RFS_C8_2	NBS	RFS	8	2	47
NBS_RFS_C8_3	NBS	RFS	8	3	47
NBS_S1_C8_1	NBS	S1	8	1	47
NBS_S1_C8_2	NBS	S1	8	2	47
NBS_S1_C8_3	NBS	S1	8	3	47
NSS_C0	NSS	-	0	-	91
NSS_RFS_C8_1	NSS	RFS	8	1	47
NSS_RFS_C8_2	NSS	RFS	8	2	47
NSS_RFS_C8_3	NSS	RFS	8	3	47
NSS_S1_C8_1	NSS	S1	8	1	46
NSS_S1_C8_2	NSS	S1	8	2	46
NSS_S1_C8_3	NSS	S1	8	3	47

NBS = Nebraska B Synthetic, NSS = Nebraska Stiff Stalk Synthetic
RFS = Reciprocal full-sib selection, S1 = S1-progeny selection

Table 3.2. Site depth, individual depth, number of polymorphic markers and gene diversity (D) in each cycle 0 and cycle 8 population. Polymorphic markers were defined as markers with a minor allele frequency greater than 2 % in cycle 0 and greater than 4 % in cycle 8.

Population	Site depth	Individual depth	Polymorphic sites	D
NBS_C0	4.148	4.148	228 502	0.132
NBS_RFS_C8_1	5.137	5.193	138 775	0.099
NBS_RFS_C8_2	4.781	4.668	106 315	0.089
NBS_RFS_C8_3	5.258	5.456	195 285	0.105
NBS_S1_C8_1	5.103	5.225	159 183	0.100
NBS_S1_C8_2	4.981	4.888	136 438	0.090
NBS_S1_C8_3	5.344	5.484	214 861	0.109
NSS_C0	4.198	4.840	276 470	0.142
NSS_RFS_C8_1	5.014	5.124	170 603	0.108
NSS_RFS_C8_2	5.100	5.025	191 130	0.113
NSS_RFS_C8_3	4.763	4.794	141 191	0.106
NSS_S1_C8_1	5.274	5.287	213 890	0.117
NSS_S1_C8_2	5.330	5.387	231 911	0.123
NSS_S1_C8_3	5.116	5.258	191 096	0.116
All Populations	3.955	4.483	312 102	--
Std. Deviation	2.403	1.340	--	--
Std. Error	--	--	--	0.0003

Table 3.3. Modified Roger's Distance (MRD) with standard errors between replicate populations comprising the UNL-RpRS program. Bolded numbers represent MRD between complementary populations. MRD was computed by selecting a random set of 10,000 markers from the total set of 312,102 markers.

	NSS						
Populations	C0	RFS_C8_1	RFS_C8_2	RFS_C8_3	S1_C8_1	S1_C8_2	S1_C8_3
NBS_C0	0.117 ± 0.002	0.167 ± 0.002	0.155 ± 0.002	0.158 ± 0.002	0.162 ± 0.002	0.146 ± 0.002	0.166 ± 0.002
NBS_RFS_C8_1	0.160 ± 0.002	0.192 ± 0.003	0.178 ± 0.003	0.185 ± 0.003	0.184 ± 0.003	0.170 ± 0.002	0.186 ± 0.003
NBS_RFS_C8_2	0.160 ± 0.002	0.194 ± 0.003	0.178 ± 0.003	0.185 ± 0.003	0.189 ± 0.003	0.171 ± 0.003	0.189 ± 0.003
NBS_RFS_C8_3	0.166 ± 0.002	0.199 ± 0.003	0.183 ± 0.003	0.191 ± 0.003	0.187 ± 0.003	0.176 ± 0.003	0.192 ± 0.003
NBS_S1_C8_1	0.165 ± 0.002	0.196 ± 0.003	0.179 ± 0.003	0.189 ± 0.003	0.190 ± 0.003	0.173 ± 0.003	0.192 ± 0.003
NBS_S1_C8_2	0.167 ± 0.003	0.198 ± 0.003	0.183 ± 0.003	0.192 ± 0.003	0.192 ± 0.003	0.170 ± 0.003	0.193 ± 0.003
NBS_S1_C8_3	0.184 ± 0.003	0.212 ± 0.003	0.197 ± 0.003	0.209 ± 0.003	0.203 ± 0.003	0.184 ± 0.003	0.206 ± 0.003

3.9 Figures

Figure 3.1. Marker summary statistics. A) SNP density in 500 Kb bins, B) Genome-wide tag depth smoothed using a Gaussian kernel (windowsize=500 Kb, stepsize=2Kb), C) Genotype quality smoothed using a Gaussian kernel (windowsize=500 Kb, stepsize=2Kb).

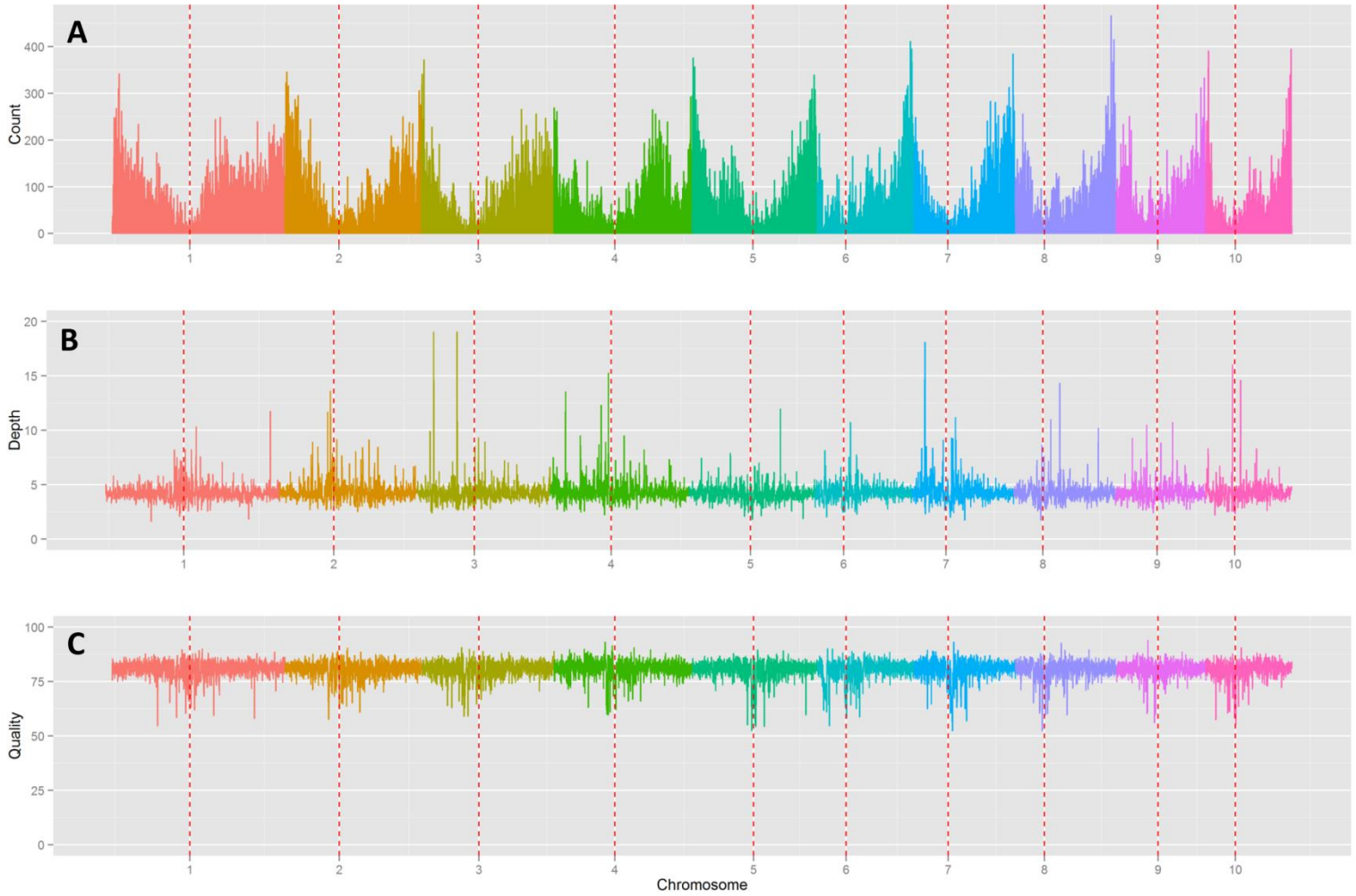

A
Count
400
300
200
100
0
1
2
3
4
5
6
7
8
9
10
B
Depth
20
15
10
5
0
C
Quality
100
75
50
25
0
Chromosome

Figure 3.2. Kernel smoothed Fst between the NBS_C0 and NBS_RFS_C8 populations for each locus. A) Fst between NBS_C0 and NBS_RFS_C8_1, B) Fst between NBS_C0 and NBS_RFS_C8_2, C) Fst between NBS_C0 and NBS_RFS_C8_3, D) Fst between NBS_C0 and NBS_RFS_C8_pooled.

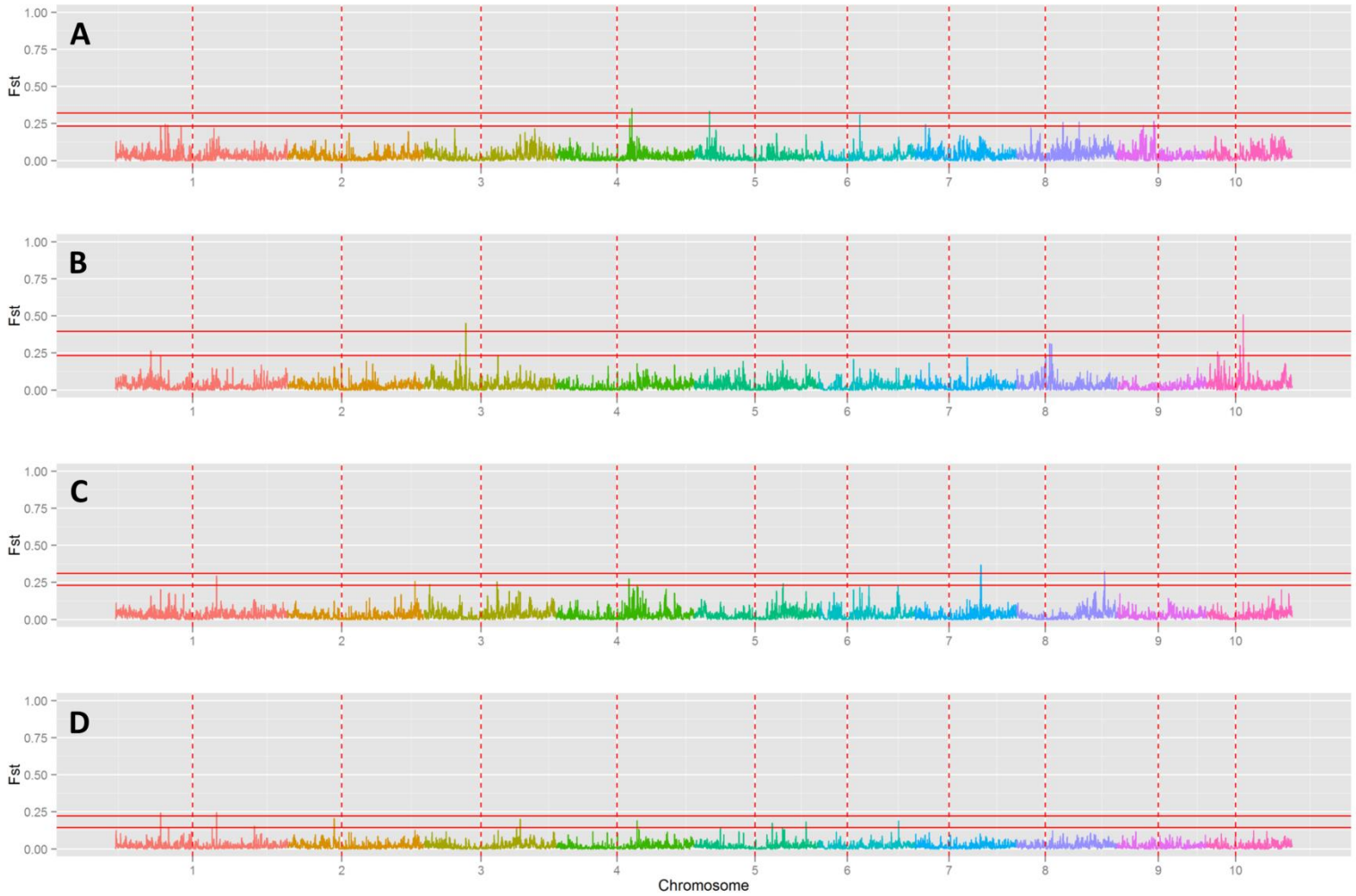

A
B
C
D
Fst
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 3.3. Kernel smoothed Fst between the NBS_C0 and NBS_S1_C8 populations for each locus. A) Fst between NBS_C0 and NBS_S1_C8_1, B) Fst between NBS_C0 and NBS_S1_C8_2, C) Fst between NBS_C0 and NBS_S1_C8_3, D) Fst between NBS_C0 and NBS_S1_C8_pooled.

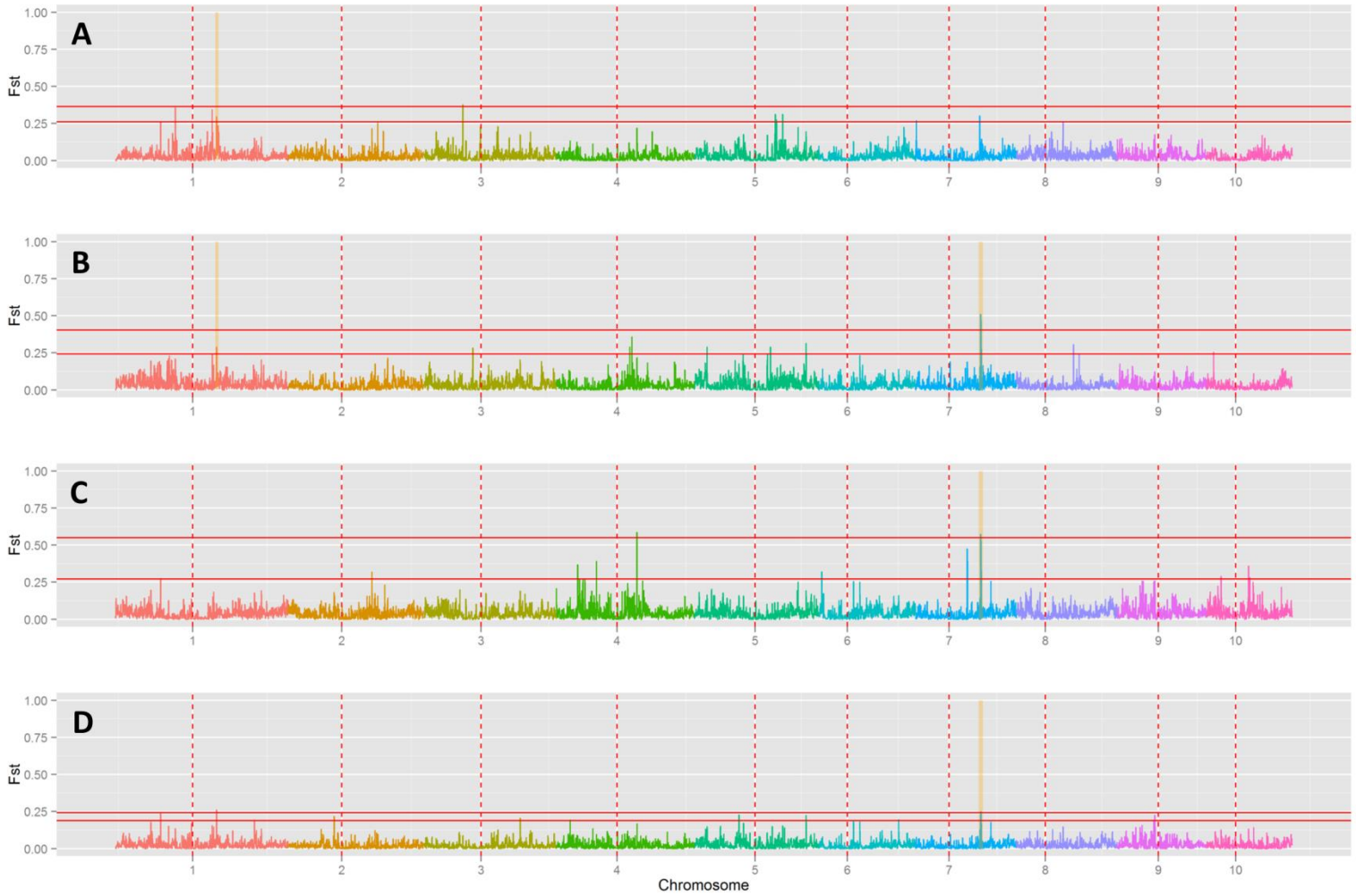

A
B
C
D
Fst
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 3.4. Kernel smoothed Fst between the NSS_C0 and NSS_RFS_C8 populations for each locus. A) Fst between NSS_C0 and NSS_RFS_C8_1, B) Fst between NSS_C0 and NSS_RFS_C8_2, C) Fst between NSS_C0 and NSS_RFS_C8_3, D) Fst between NSS_C0 and NSS_RFS_C8_pooled.

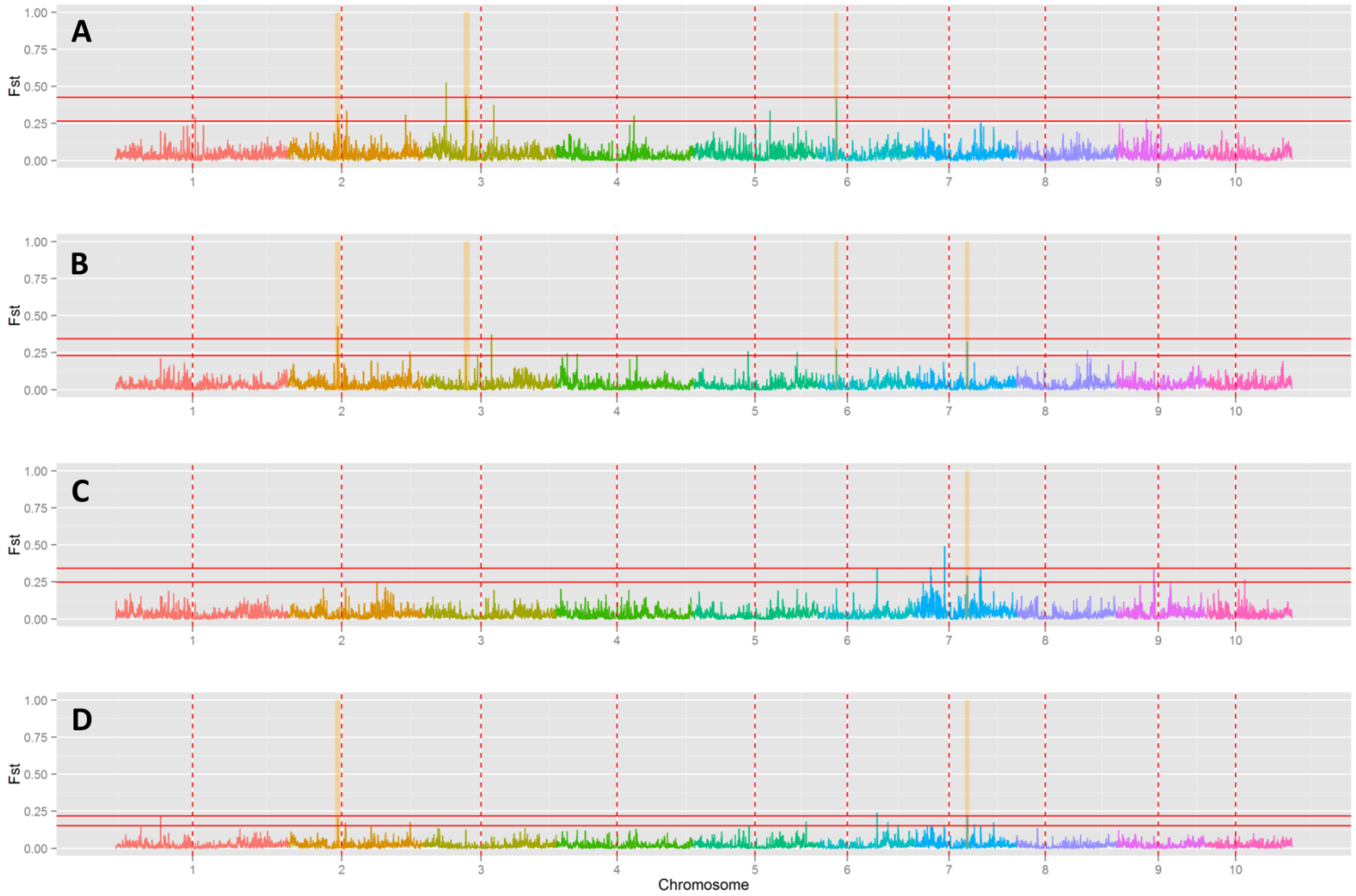

A
B
C
D
Fst
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 3.5. Kernel smoothed Fst between the NSS_C0 and NSS_S1_C8 populations for each locus. A) Fst between NSS_C0 and NSS_S1_C8_1, B) Fst between NSS_C0 and NSS_S1_C8_2, C) Fst between NSS_C0 and NSS_S1_C8_3, D) Fst between NSS_C0 and NSS_S1_C8_pooled.

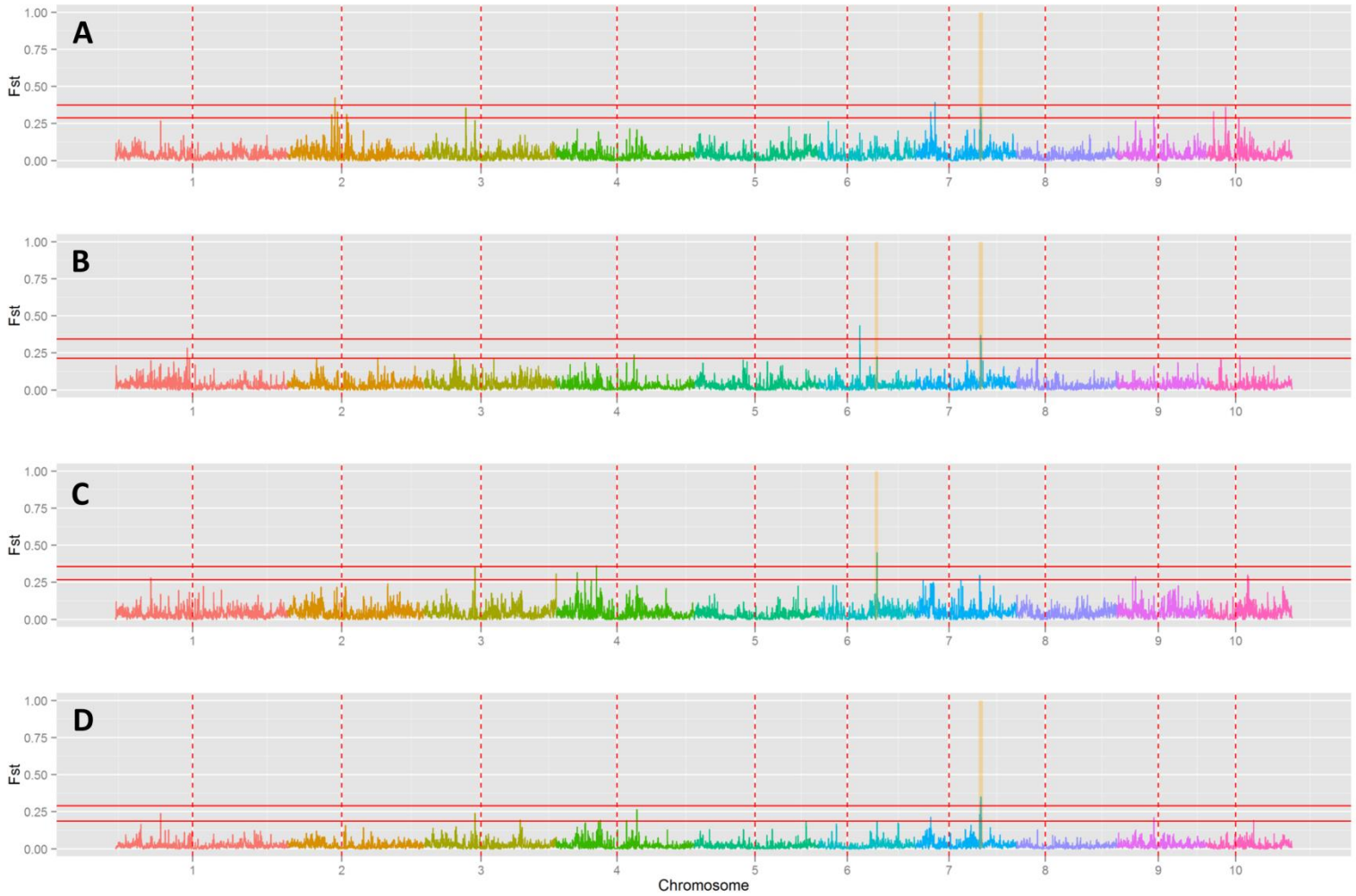

A
B
C
D
Fst
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 3.6. Eigen vector plot of the UNL-RpRS program. The first and second Eigen vectors explain 4.78 and 1.51 % of the variation, respectively.

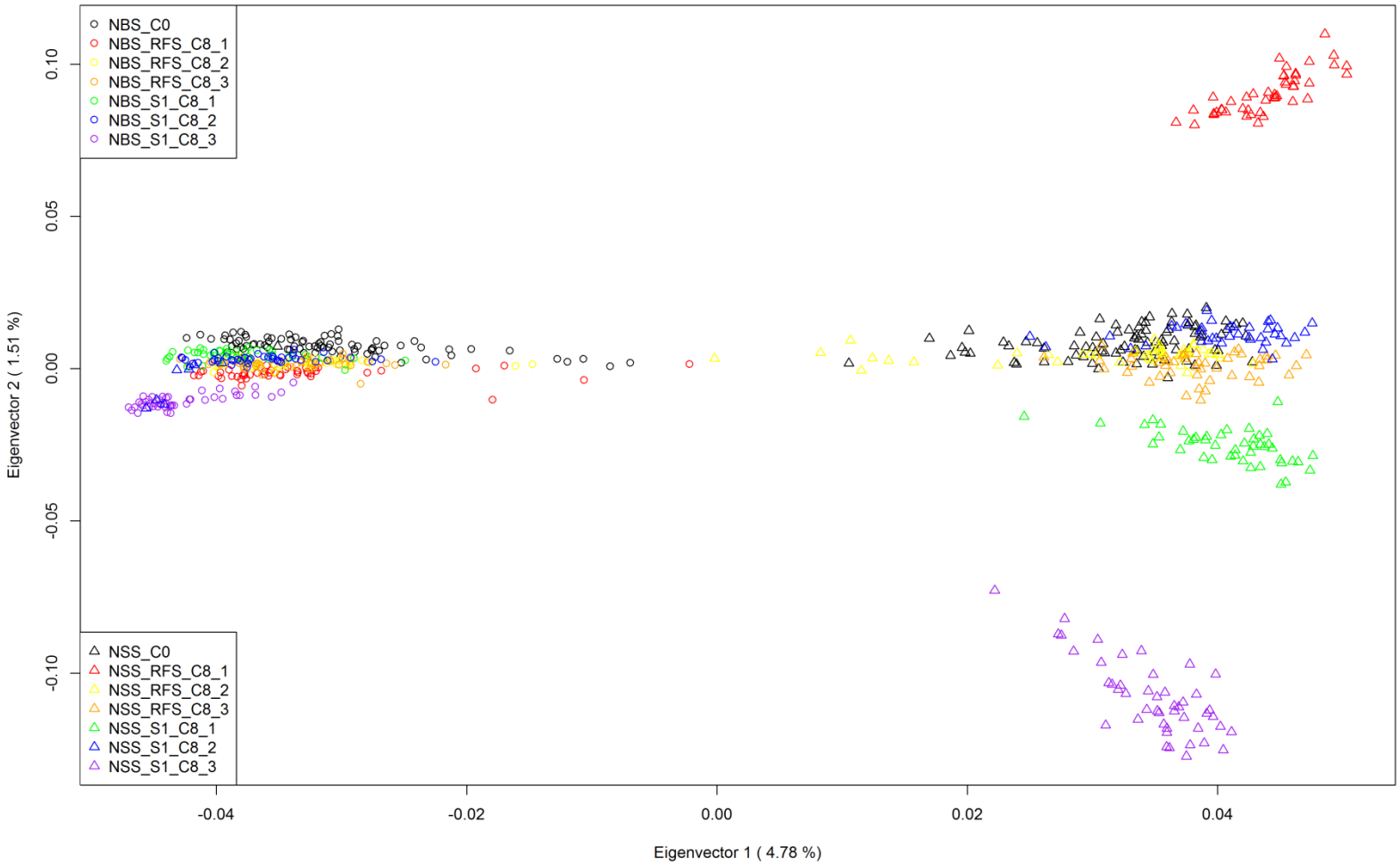

NBS_C0
NBS_RFS_C8_1
NBS_RFS_C8_2
NBS_RFS_C8_3
NBS_S1_C8_1
NBS_S1_C8_2
NBS_S1_C8_3
NSS_C0
NSS_RFS_C8_1
NSS_RFS_C8_2
NSS_RFS_C8_3
NSS_S1_C8_1
NSS_S1_C8_2
NSS_S1_C8_3
0.10
0.05
0.00
-0.05
-0.10
Eigenvector 2 (1.51 %)
-0.04
-0.02
0.00
0.02
0.04
Eigenvector 1 (4.78 %)

Figure 3.7. Kernel smoothed Fst between reciprocally selected RFS populations corrected for divergence already present between NBS_C0 and NSS_C0 for each locus. A) Fst between NBS_RFS_C8_1 and NSS_RFS_C8_1, B) Fst between NBS_RFS_C8_2 and NSS_RFS_C8_2, C) Fst between NBS_RFS_C8_3 and NSS_RFS_C8_3 and D) Fst between NBS_RFS_C8_pooled and NSS_RFS_C8_pooled.

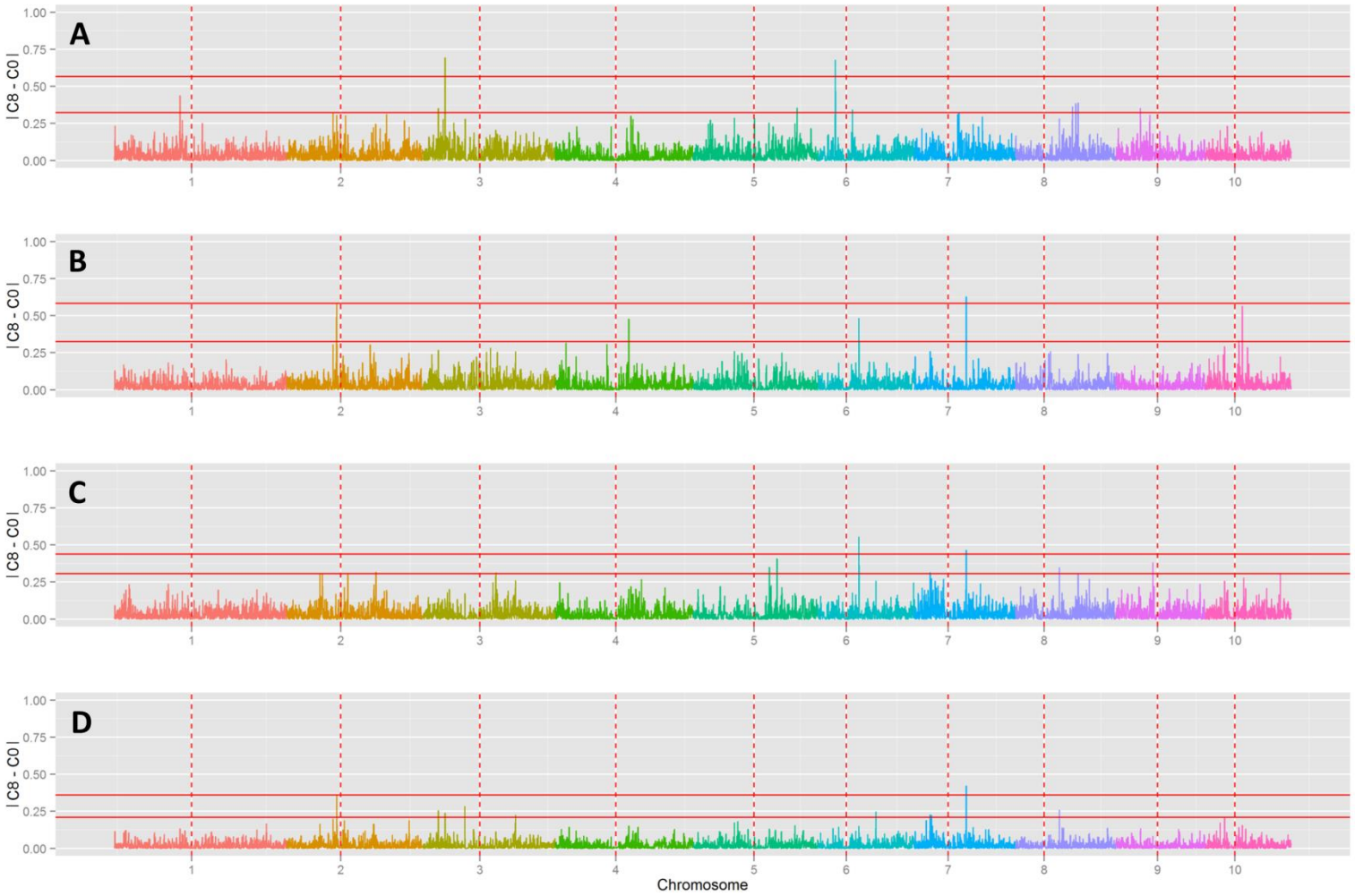
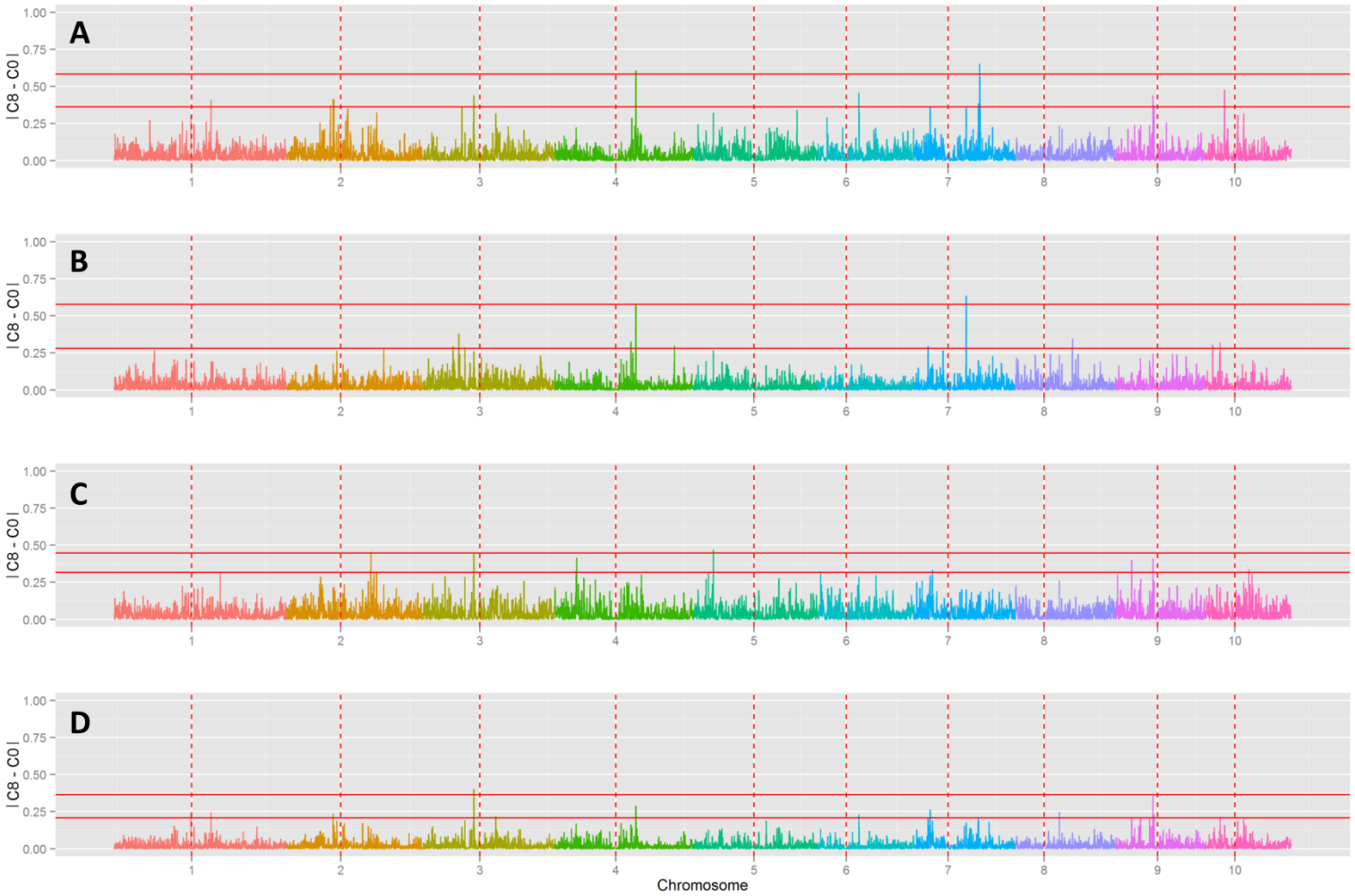

A
B
C
D
|C8 - C0|
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 3.8. Kernel smoothed Fst between reciprocally selected S1 populations corrected for divergence already present between NBS_C0 and NSS_C0 for each locus. A) Fst between NBS_S1_C8_1 and NSS_S1_C8_1, B) Fst between NBS_S1_C8_2 and NSS_S1_C8_2, C) Fst between NBS_S1_C8_3 and NSS_S1_C8_3, D) Fst between NBS_S1_C8_pooled and NSS_S1_C8_1_pooled.


A
B
C
D
| C8 - C0 |
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
7
8
9
10
Chromosome

4 CHAPTER 4 POPULATION STRUCTURE AND IDENTIFICATION OF GENOMIC REGIONS PUTATIVELY ASSOCIATED WITH SELECTION IN A LARGE SET OF INBRED LINES

4.1 Abstract

The previous century of maize breeding saw the formation and development of heterotic groups and heterotic patterns in maize. Examining genetic changes which accompanied selection for hybrid performance have revealed very pronounced changes in diversity and population structure similar to trends seen in reciprocal recurrent selection programs. Specifically, a reduction in genetic diversity accompanied by the development of pronounced population structure in time. Previous work by other researchers has provided access to a large set of densely genotyped inbred lines representative of inbred lines popular throughout maize breeding history. Our objectives were to 1) evaluate and describe the genetic structure in this large sample of maize inbred lines delineated by four time points and, 2) perform a genome-wide scan for localized signatures of selection left by selection for hybrid performance between heterotic groups. We accomplished this by grouping inbred lines into eras based on date of release, accompanied by genetic grouping done with eigenanalysis. A search for localized signatures of selection was performed using the Fst statistic within eras, using the defined genetic groupings as subpopulations. We found that, consistent with current research, the canonical story suggesting the importance of the Reid-Lancaster heterotic pattern does not hold up, and

our data support the notion heterotic groups have developed empirically through trial and error over time. Consistent with current research, a small set of prominent inbred lines came to dominate the ancestry of newly developed inbred lines within heterotic groups. A search for localized signatures of selection revealed little evidence for strong selection, and a small set of windows show evidence for a mode of selection maintaining diversity in these populations.

4.2 Introduction

Since Shull's (1908) discovery of the single-cross hybrid, maize breeders have aggressively pursued the goal of maximizing hybrid performance. As often is the case with scientific endeavors, the path to economically viable single-cross hybrids was not linear. Early inbred lines developed from open-pollinated landrace populations suffered badly from inbreeding depression, with Raymond Baker on record (Baker, 1984; Tracy and Chandler, 2006) commenting that "Just keeping those early inbreds from open-pollinated corn alive was an art." The issue of severe inbreeding depression was circumvented in two ways. First, Jones (1918) developed the concept of the double-cross hybrid, which used F1 parents in production fields, making hybrid seed production economically viable. Secondly, inbreeding depression is a heritable trait (Pray and Goodnight, 1995; Edwards and Lamkey, 2002). As superior inbred lines were identified they were recycled by using them as parents for the next "cycle" of selection (Hallauer, 1988; Tracy and Chandler, 2006), reducing inbreeding depression with time (Duvick, et al., 2004). Through the 1950s and 1960s, inbred performance eventually improved to the

point where single-cross hybrids were economically viable, and by 1960 nearly all the maize acreage in Iowa was planted to single-cross hybrids (Hallauer, 1988).

The very act of inbred improvement gave rise to one very critical issue with respect to developing hybrids, which inbred lines make the best hybrids? This question, in turn, gave rise to a second issue, how should inbred lines be grouped to maximize the probability of developing superior hybrids? To address this second issue, early corn breeders met at the annual Corn Improvement Conference and developed a subcommittee titled "Committee on Grouping Inbred Lines for Breeding Purposes" (Anon, 1949). Initially, inbred lines were split into A and B groups at random (Anon, 1949; Tracy and Chandler, 2006) and heterotic groups as we know them today were formed empirically as a result of trial and error (Tracy and Chandler, 2006).

Melchinger and Gumber (1998) define a heterotic group as "… a group of related or unrelated genotypes from the same or different populations, which display similar combining ability and heterotic response when crossed with genotypes from other genetically distinct germplasm groups." Heterotic groups are necessarily dynamic over time because breeding within heterotic groups will, itself, give rise to improved inbred lines. Extensive research has characterized the genetic origins of heterotic groups and changes in pedigree backgrounds through time (Troyer, 1999; Smith et al., 2004; Mikel and Dudley, 2006) and more recently, molecular marker research has revealed changes in genetic diversity and population structure (Duvick et al., 2004; Ho et al., 2005; Nelson et al., 2008).

The system for developing inbred lines and testing hybrids previously described very closely resembles a reciprocal full-sib (RFS) selection program (Duvick et al., 2004). Genetic changes as a result of reciprocal recurrent selection (RRS; Comstock et al., 1949), of which RFS selection is a derivative, have been closely documented. Two trends are typically observed: a reduction in genetic diversity in time and a corresponding increase in population structure (Hinze et al., 2005; Lamkey and Lorenz, 2014). These same trends have been observed in the inbred-hybrid system of maize breeding (Duvick et al., 2004; van Heerwaarden et al., 2012).

While changes in genetic diversity within and among heterotic groups has been well characterized, relatively few studies have searched for localized signatures of selection. van Heerwaarden et al. (2012) used the Illumina MaizeSNP50 beadchip (Ganal et al., 2011) coupled with a novel association mapping method and correlated allele frequency changes with time, finding that selection is relatively modest and consistent with a model of adaptation from standing genetic variation. Other studies have examined diverse sets of inbred lines and reached similar conclusions regarding the impact selection has in modern breeding (Hufford et al., 2012).

Previous studies (Duvick et al., 2004; Ho et al., 2005; Nelson et al., 2008; van Heerwaarden et al., 2012) selected inbred lines which were popular in hybrids and made significant contributions to the germplasm base. Here, we collected a large set of inbred lines identified from the North Central Corn Breeding Committee meeting minutes, as well as inbred lines previously used by van Heerwaarden et al. (2012). All lines were

previously genotyped using genotype-by-sequencing (GBS), which generated a dense set of SNPs. Our objectives were to 1) evaluate and describe the genetic structure in this large sample of maize inbred lines delineated by four time points and, 2) perform a genome-wide scan for localized signatures of selection left by selection for hybrid performance between heterotic groups.

4.3 Materials and Methods

4.3.1 Selection of inbred lines

The set of inbred lines used in this study were selected by pooling information from two sources. First, van Heerwaarden et al. (2012) conducted a study using a set of 301 inbred lines and 99 landrace varieties. Only the inbreds were included. Generally, these lines were relatively new inbred lines, with a proportionally large collection of inbred lines being ex-plant variety protection (ex-PVP). Second, the NCCC167 meeting minutes, formerly the North Central Corn Breeding Technical Committee, were reviewed from 1949 (Anon, 1949) to 1987 (Anon, 1987). Periodically, the sub-committee for grouping inbred lines for breeding purposes would update a list with new inbred lines and drop old inbred lines. This list represents inbred lines used actively during the past century of maize breeding. The inbred lines from the van Heerwaarden et al (2012) study and the NCCC167 list were compiled into a single set. Generally, the collection of inbred lines from the NCCC167 list consisted of older inbred lines, while the van Heerwaarden et al. (2012) set consisted of newer inbred lines.

The inbred list developed above was then used to select a subset of inbred lines genotyped by Romay et al. (2013) from the Ames Plant Introduction Station (Ames, IA). When possible, a GBS Identifier was attached to each inbred. The GBS Identifier is the name of the inbred line as it appears in the genotype files downloaded from www.panzea.org. This list was checked for duplicate entries by searching for identical plant introduction numbers, inbred names, and GBS identifiers. Finally, the list was inspected manually for any duplicated lines.

4.3.2 Genotype data and data filtering

The GBS data for this experiment was originally generated by Romay et al. (2013) and is available for download from www.panzea.org. The retained GBS Identifiers were used to subset the original chromosome files. Specifically, the partially imputed genotype data was downloaded. Genotype data was filtered for no more than 10 % missing data and minor allele frequency of 1 %. In total, 570 inbred lines and 529 253 SNP markers were retained for analysis.

4.3.3 Era assignment

All retained inbred lines that had a GBS Identifier were assigned to a decade based on date of release. Decades ranged from 1930 to 1980. Approximately, 20 inbred lines were released in the 1990s, but these were grouped with the 1980's inbreds because of small sample size. Gerdes et al. (1993) was used to assign decades based on date of publication. In several instances, known historical inbred lines had very recent references or a different reference was provided (e.g. recent personal communication reference or a

circular reference to another inbred). In these instances, a secondary reference was sought. Secondary references included Henderson (1976, 1980, 1984) and as a last resort, the date of "first appearance" in the NCCC167 meeting minutes was used. The data of first appearance was the first time an inbred line appeared in the subcommittee for Grouping Inbred Lines for Breeding Purposes meeting minutes. If a date could not be determined the line was dropped. Decade assignment is being used as a surrogate for the time frame in which the inbred was used for breeding purposes. Assigning time points in this way is not always representative of when an inbred was actually used in breeding programs or commercial hybrid production, but provides a reasonable, and in several cases the only, alternative for determining chronology.

Clustering within some decades was poor due to lack of structure. This was particularly true in the early decades (i.e. 1930 and 1940; data not shown). To circumvent this problem, four Eras were created. Era 1 contained inbreds released in 1930 and 1940, Era 2 contained inbreds released in 1950 and 1960, Era 3 contained inbred released in the 1970's, and Era 4 contained inbreds released in the 1980's (Table 1). Broadly, these Eras correspond to the type of hybrid the inbred was used in. Era 1 represents the double cross era, Era 2 represents the transition from three-way cross to single cross, Era 3 represents public single-cross lines and Era 4 represents the ex-PVP era, which can be thought of as an "advanced" single-cross era.

4.3.4 Genetic group assignment

Inbred lines were assigned to genetic groups within Eras. The methodology of Patterson et al. (2006) was implemented. The conventional SS-NSS grouping was not used because prior to the 1960's our data suggests a clear SS-NSS trend was not yet detectable. Patterson et al. (2006) presented an algorithm that identifies population structure based on eigenanalysis. Specifically, the algorithm identifies a set of significant eigenvalues by comparing them to a Tracy-Widom (TW) distribution. For completeness, the algorithm proposed by Patterson et al. (2006) will be outlined here.

Prior to running the algorithm, marker data were recoded so that x {0,1,2}, where x represents the number of copies of the major allele an individual carries. Missing data was imputed using naïve imputation. The numeric marker matrix was then normalized by

$$M(i,j) = \frac{C(i,j) - u(j)}{\sqrt{p(j)(1-p(j))}}$$

where $C(i,j)$ is the value (0, 1, 2, or the column mean in the case of missing data) of the *ith* row and *jth* column of the genotype matrix, $u(j)$ is the mean of the *jth* column, $p(j)$ is the frequency of the major allele for the *jth* column, and $M(i,j)$ is the normalized genotype value for the *ith* row and the *jth* column. M is now the normalized genotype matrix.

The covariance matrix was then computed as

$$X = MM`$$

where X is the normalized covariance matrix.

Eigenanalysis is then performed on the covariance matrix X. In order to compare eigenvalues to the TW distribution they first have to be normalized. Let L_i be the *ith* eigenvalue, then

$$L_i = \frac{(m`)\lambda_i}{\sum_{i=1}^{m`} \lambda_i}$$

where $m` = m - 1$ and λ_i is the *ith* eigenvalue.

Then the normalized eigenvalue is

$$x = \frac{L_i - u(m,n)}{\sigma(m,n)}$$

where $u(m,n) = \frac{\left(\sqrt{n`-1}+\sqrt{m}\right)^2}{n`}$ and $\sigma(m,n) = \frac{\left(\sqrt{n`-1}+\sqrt{m}\right)}{n`}\left(\frac{1}{\sqrt{n`-1}} + \frac{1}{\sqrt{m}}\right)^{1/3}$ and $m$ is the number of markers, $n`$ is the effective number of markers, which is a statistical value *not* the actual number of markers (see Patterson et al., 2006). x is approximately TW (Patterson et al., 2006). The RMTstat package (Johnstone et al., 2009) in R was used to generate a TW distribution. Significant p-values ($p < 0.05$) were used to identify a set of eigenvalues for each decade. The number of eigenvectors used was determined by the number of significant eigenvalues.

Euclidean distances among individuals were computed within each Era based on the number of eigenvectors identified from their corresponding eigenvalues. Ward

clustering was then performed and genetic groups were selected. Once groups were identified, group membership was retained in the "Group" column of Supplementary Dataset 1.

4.3.5 Summary statistics

Allele frequencies were estimated using maximum likelihood (Weir, 1996). Fst (Wright, 1978) was computed following Weir and Cockerham (1984). This estimator corrects for a small number of populations and unequal sample sizes, both of which are issues in this study (Weir and Cockerham, 1984). Fst ranges between 0 and 1, but, because Fst was estimated, the value can be negative or undefined. Negative values were retained, while undefined values were set to zero.

Genetic diversity Weir (1996) was captured in each Era and computed as

$$D = 1 - \sum_{i=1}^{k} p_i^2$$

where k denotes the number of alleles and p_i^2 is the square of the *ith* allele frequency. For biallelic loci, gene diversity ranges between 0 and 0.5.

4.3.6 Data smoothing

Fst values were smoothed with a Gaussian kernel as described in Hohenlohe et al. (2010) and computed as

$$f(x) = exp\left(\frac{-(x-c)}{2\sigma^2}\right)$$

where x is the physical coordinates of the SNP in question, c is physical coordinates of the SNP in the center of the window and σ is the size of window in base pairs. Thus, Fst values within a window are weighted according to their proximity to the SNP in the center of the window. As SNPs become distal to the center SNP their weight decreases.

The window size and step size were 400 Kb and 100 Kb, respectively. Window size was selected by computing several different window sizes and finding a window size large enough to smooth the data, but not so large as to homogenize the genetic signal. Since window size is a property of the physical map, sometimes gaps in the genome cause windows to be excessively large, creating uncertainty about the estimates of the kernel. To avoid these windows, all windows falling outside of two standard deviations were dropped. This corresponded to windows larger than 442 615 bp and smaller than 357 384 bp.

4.3.7 Identification of putative selection candidates and underlying gene models

An outlier test was applied to distributions of Fst and gene diversity to identify windows that are putative candidates for selection. The outlier test was performed at the 99.9 % and 99.99 % levels to identify strong and very strong evidence for selection, respectively. The outlier test has been commonly used in the literature (Black et al., 2001; Akey, 2009; Beissinger et al., 2013). Gene models that might be associated with selection were identified in each 400 Kb window that passed the outlier test.

4.4 Results

4.4.1 Population structure

The list of inbred lines was selected to represent the full scope of maize breeding history from 1930 through the 1980s. To this end we succeeded in developing a comprehensive list of 570 public and exPVP inbred lines, and captured the genetic structure expected as a result of the inbred-hybrid system of maize breeding. The observed structure corresponded to the SS and NSS heterotic groups (Figure 1). The genetic structure and important inbred lines will be discussed for each Era in turn.

Era one was split into two genetic groups (Figure 2). Group one tended two have fewer lines derived from open-pollinated landrace populations and contained more crosses among early inbred lines (Figure 2). Ancestry traces back to various yellow dent varieties including Osterland Yellow Dent and Funk Yellow Dent. C103, one parent of Mo17 was included in Group one, along with Wf9, and several seminal Iodent lines including I205. Group two contained several inbred lines developed directly from open-pollinated landrace populations or synthetics including Golden Glow, Silver King, Pride of Saline, and Minnesota No. 13. Era one represents the origins of the modern heterotic pattern observed today.

Era 2 characterizes inbred lines developed in the 1950's and 1960's (Figure 3). The majority of inbred lines fall into genetic group 1, which represents a largely undifferentiated mass. Genetic group 1 contains major contributions from the Illinois and

Minnesota breeding programs. The Illinois program contributed several lines originating from Snelling Corn Borer Synthetic and Illinois Synthetic in addition to multiple *br2* (brachytic2) recoveries of notable Era 1 lines such as C103 as well as Era 2 lines (CI21E). Genetic groups 2 and 3 correspond to early SS and NSS heterotic groups, respectively with B14 and A630 characterizing each group (Figure 3). Genetic Group 2 corresponded to the SS heterotic group and is characterized by B14 and, several Minnesota inbred lines (A632, A634, A639, A641) that share B14 as a parent. Interestingly Genetic Group 2 includes a subgroup which consists of a set of inbred lines characterized by M14, which is the cross between two lines (BR10 x R8) developed from Texas Surecropper (R8) and Funk Yellow Dent (BR10). The NSS heterotic group is nearly entirely characterized by the Era 1 line Wf9 with almost every inbred having Wf9 in its pedigree. Wf9 was developed in Indiana from Reid Yellow Dent.

The modern heterotic pattern observed today was a product of maize breeders running pedigree selection within heterotic groups and a hybrid testing program between heterotic groups. In contrast to RRS programs, which use genetically closed populations, at any time breeders could develop new inbred lines from landrace populations, older inbreds, hybrids, or conversions of older inbreds. Developing inbred lines in advanced eras from landrace populations or hybrids will make an inbred appear undifferentiated with respect to its contemporaries developed from recycled inbred lines. This trend can be observed in each of eras two, three and four (Figures 3-5).

Era 3 characterizes lines developed in the 1970's and contained five genetic groups (Figure 4). Genetic group 1 contains undifferentiated germplasm such as gametophytic male sterile lines (W series from Missouri), conversions of several older inbred lines (Missouri 300 series) for the *ae* gene (amylose extender), as well as other resistance or enhancement traits. Genetic group 2 corresponds to the SS heterotic group. Interestingly, this group receives a large portion of its parentage from B14 and the BSSS synthetic population. Inbred B68 characterized genetic group 2 and is highly related to B14 (Gerdes et al., 1993). Genetic group 3 corresponded to the NSS heterotic group and is characterized by Pa762 which has Oh43 in its parentage (Gerdes et al., 1993). The previous Era received a very large contribution of its parentage from Wf9. In Era 3, C103, and Oh43 were very prominent NSS inbred lines contributing to parentage with the rest of the heterotic group being mixed. Group 4 from Era 3 is composed of five inbred lines all with inbred 38-11 as a parent. 38-11 is a 1930s inbred developed in Indiana (Gerdes et al., 1993). In the case of Group 5, three of the inbred lines result from a cross between Mo22 x Wf9 or the reciprocal cross. The remaining inbred is selfed out of a landrace called Laguna, which originated in Mexico (Smith et al., 2004).

Era 4 characterizes inbred lines developed in the 1980s and contained five genetic groups (Figure 5). In contrast to genetic groups three through five, genetic groups one and two were largely composed of exPVP inbred lines. In Eras two and three the Minnesota, Indiana, and Illinois breeding programs made fairly large contributions to the genetic structure of the important heterotic groups. In Era 3 the Iowa breeding program

made significant contributions to the formation of the SS heterotic group. By Era 4, the contribution of Illinois, and to a lesser degree Indiana had declined. Breeding programs from North Dakota made contributions to this group. The 1980s was a pivotal time in the corn breeding community. The last NCCC167 sub-committee for grouping germplasm for breeding purposes disbanded in 1987 (Anon, 1987), and the private sector in corn breeding had significantly developed. This can be seen by examining the contributions to genetic groups one through four. Large contributions from Pioneer Hi-Bred, Holdens, and several other private companies dominate these groups. Since the germplasm was proprietary and was mostly developed from coded lines, (i.e. inbred lines given an official designation by companies), interpreting the genetic contributions from pedigrees is difficult. However, some conclusions can be drawn. Genetic groups 1 and 2 are the most polarized in Figure 5, and correspond to the NSS and SS heterotic groups, respectively. Genetic group 1 is characterized by LH51, a Holdens inbred that is highly related to Mo17. Genetic group 2 is characterized by inbred LH74, which is related to B14 and B73. Genetic group 3 occupies the center of Figure 5 and is a mixture of Pioneer Hi-Bred inbred lines and public inbred lines. Genetic group 4 is a relatively small subset of almost exclusively Pioneer Hi-Bred lines.

4.4.2 Genome-wide patterns of genetic differentiation

The second objective of this study was to identify localized signatures of selection on the genome. Although genetic differentiation increased from era one to era four (Figure 6), we observed a reduced degree of differentiation among Eras compared with

van Heerwaarden et al. (2012). Certainly, the reduced degree of differentiation was due to the undifferentiated inbred lines present in each of Eras two, three and four.

Genome-wide patterns of kernel-smoothed Fst values revealed 8 regions which passed the outlier test at 99.99 % (Figure 7). In Era 1, chromosome 9:104031536-104434789 and chromosome 10:43503486-43876703 were identified (Figure 7A). In Era 2, two regions were located on chromosome 9, which were very close, located at 44073006-44452360 bp and 45663639-46055306 bp (Figure 7B). In Era 3, two windows were identified at unique positions on chromosome 7 at 78812761-79211426 bp and 102869411-103277727 bp (Figure 7B). In Era 4, chromosome 1:123318407-123744703 bp and chromosome 6:21426754-21855036 bp were identified (Figure 7D). This scan implicated 15 gene models, eight of which did not have a characterized function. The remaining seven had various descriptions, of note was gene model GRMZM2G476848 identified on chromosome 9 in Era 2 and is described as a pollen specific arabinogalacta protein and may be associated with cold tolerance (Darvill et al., 1994).

Kernel smoothed estimates of gene diversity were also examined (Figure 8). It appeared that some regions in each of the eras showed relatively high levels of diversity when compared to an outlier test at 99.9 %. These regions represent regions which might candidates for balancing selection. If these regions were under balancing selection it would be reasonable to think that they might be consistent across eras because balancing selection acts to maintain variation instead of fix an allele (Hartl and Clark, 2006). The intersections between adjacent eras were examined for a set of loci showing this signature

of selection and eight regions were putatively identified. The regions were identified on chromosomes one, three, six (two adjacent regions), eight (two adjacent regions), and 10 (Figure 8A-D). Each region displayed relatively low values of Fst with elevated gene diversity (Table 2), which conforms to a classic signature for regions under balancing selection. In the case of the region on chromosome 1, gene diversity is held relatively constant across eras (0.358, 0.383, 0.373, 0.378), but Fst increased from 0.003 in Era 1 to 0.256 in Era 4 (Table 2). Similarly, a region on chromosome 10 shows the same pattern (Table 2). The remaining regions all show similar patterns, but to a much reduced degree than that observed for the regions on chromosomes one and 10. This subset of genes implicated 13 gene models eight of which were not characterized. The remaining five gene models had various functions. One interesting gene model (GRMZM2G025105) was identified in Era 2 on chromosome 6. This gene model is a polygalacturonase inhibitor, which is an enzyme produced by bacterial and fungal pathogens. Inhibition of this enzyme may be correlated with resistance to various pathogens.

4.5 Discussion

Here we sought to characterize the genetic structure of a large collection of inbred lines revealed by temporal grouping and to identify regions under selection in the genome. The discussion will be split into two sections focusing on population structure and potential selection candidates. Each section will be discussed in turn followed by conclusions.

4.5.1 Population structure

As the canonical story of the formation of modern heterotic groups (Tracy and Chandler, 2006) goes, the Southern Dents (Brown and Anderson, 1948) and Northern Flints (Brown and Anderson, 1947) came to America experiencing very different environmental conditions and could be considered separate species (Anderson and Brown, 1952; Labate et al., 2003). The Southern Dents and Northern Flints were crossed both accidently then experimentally to eventually form a new race of corn termed Corn Belt Dent (Tracy and Chandler, 2006). Corn Belt Dent, in turn, gave rise to famous open-pollinated varieties like Reid and Lancaster amongst many others (Labate et al., 2003). The Reid open-pollinated variety developed in Illinois and Iowa resembled Southern Dent maize. The Lancaster variety developed in Pennsylvania resembled Northern Flint maize (Tracy and Chandler, 2006). To finish the canonical story, early corn breeders found exceptional combining ability between the Reid-Lancaster heterotic pattern and attributed this to the geographic distance separating them. Tracy and Chandler (2006) go on to show that the canonical story did not hold up to a rigorous historical analysis, and more importantly, the foundations of the modern heterotic groups were started by random designations of inbred lines to arbitrary groups (Anon, 1949) and developed empirically through trial and error.

Gerdes and Tracy (1993) show that the Lancaster Surecrop heterotic group is represented in modern maize germplasm, almost in its entirety, by Oh43 and Mo17, which are themselves only 50 % Lancaster (Gerdes and Tracy, 1993; Tracy and

Chandler, 2006). Our data show that by era three, Oh43 was contained in the parentage of many inbred lines, but, C103 was still important. This suggests that at least through the 1970s the prevalence of Lancaster germplasm in breeding crosses within heterotic groups was maintained. However, by 1980 essentially all the parentage of the NSS group in era four was Mo17. Mo17 in hybrid combination with SS inbred lines suggests the importance of Reid Yellow Dent germplasm in maize from which many of the founders of BSSS were derived (Gerdes and Tracy, 1993). Taken together, our data support the results of Tracy and Chandler (2006).

Starting with Era two, B14 and Wf9, were found to be major contributors to the parentage of inbred lines characteristic of Era three. In Era three, B14 remained an important SS inbred line, but, RRS and inbred development in the BSSS-BSCB1 recurrent selection program produced inbred line B73. B37 also gained popularity and was developed from C0 of BSSS-BSCB1 RRS program. In the NSS germplasm group C103 and Oh43 were very prominent inbred lines, and inbred C103 gave rise to Mo17. By Era four, essentially all the NSS inbred lines observed were related to Mo17, and B73 had gained a very strong foot hold in the parentage of this Era, especially in the Holden's germplasm included in this study. The groupings for our set of lines correspond well with previous work by Liu et al. (2003) who analyzed genetic diversity in a set of 260 inbred lines and found similar groupings of SS and NSS lines as identified here. Recently, van Heerwaarden et al. (2012) analyzed the ancestral contribution of important historic inbred lines to ex-PVP lines, (Eras 1 and 3, respectively in their study) and found Wf9 and Oh43

to be significant contributors and to a lesser degree C103 and B37. Results from Mikel and Dudley (2006) corroborate well with our results suggesting the importance of Oh43, Wf9, B14, and B37 early in maize breeding history accompanied by their loss of importance through time as B73 and Mo17 gained prominence.

4.5.2 Selection candidates

In accordance with van Heerwaarden et al. (2012) we found minimal evidence for strong directional selection. van Heerwaarden et al. (2012) suggested this is consistent with a model of selection acting on standing genetic variation. Selection from standing genetic variation is consistent with a soft-sweep model of evolution (Hermisson and Pennings, 2005), whereby variation is present in the genome prior to the time selection starts. This model of adaptation can distribute favorable alleles throughout the population through recombination, which also distributes the selection signal and thereby weakening the signal. Beissinger et al. (2013) found that a majority of selection candidates identified in a mass selection program fit a soft sweep model of adaptation. Lamkey (this dissertation) also found evidence for parallel evolution, which gives support to a soft-sweep model of adaptation.

Duvick et al. (2004) suggested that corn breeders have sought to develop hybrids which maintain or increase performance across increasingly adverse environments. It is reasonable to think that selection for increased yield in adverse environments has acted on genes that regulate biotic and abiotic stress response. We identified a gene (GRMZM2G476848) that is putatively involved in cold tolerance response. Elite, modern

maize germplasm is more cold tolerant, allowing earlier planting dates for a longer growing season (Mock and Eberhart, 1972; Lauer et al., 1999; Revilla et al., 2000). A second gene (GRMZM2G025105) was identified that appears to be involved with inhibiting an enzyme from bacterial and fungal pathogens. This could be a component for success with increasingly homogenous cropping systems maintaining harmful diseases in the production environment.

Eight regions were identified as putative candidates for balancing selection. Balancing selection seeks to maintain diversity at a site under selection. Selection has occurred at two levels in maize breeding programs: 1) pedigree selection on inbred lines within heterotic groups followed by 2) selection on hybrids among heterotic groups. Maize heterotic groups show sexual dimorphism where the sexes differ within a heterotic group for morphological characters. Inbred lines developed from the SS heterotic group tend to make better females and NSS inbred lines generally tend to make better males (Tracy and Chandler, 2006). Differential selection between sexes (Hartl and Clark, 2007) can favor different genotypes between sexes and thus allow for maintenance of polymorphism at a site in the total germplasm pool. Differential selection between sexes might be a plausible model for selection when considering selection for a variety of morphological characters in breeding programs. Male inbred lines require a very different set of traits compared to female inbred lines. For instance, tassel development, tassel shedding, and tassel branch number are all characteristics that are desirable in males, but

not desired in females since a bulk of the reproductive energy should go into ear development.

Multiple modes of selection, in addition to balancing selection, can give rise to maintenance of genetic diversity at a site or appear to give rise to maintenance of diversity at a site. Selection in heterogeneous environments (Hartl and Clark, 2006), where the relative fitness of a genotype changes with environment, is one model for selection that can maintain diversity at a site. This mode of selection can maintain polymorphism in the absence of overdominance (Hartl and Clark, 2006). This is accomplished by each homozygous genotype being favored in different environmental conditions. Across environments, the heterozygote is most fit even though it is not the best genotype in any particular environment. The most tempting conclusion to draw is that, perhaps, these are regions showing overdominance (Crow, 1999). In each case, drawing strong conclusions particularly about the mode of dominance responsible for these regions is highly speculative since this study it is not designed to answer these questions.

4.6 Conclusions

Here we sought to evaluate the genetic structure of a large set of inbred lines developed as a result of selection for hybrid performance and to characterize regions which may be under selection in the genome. We evaluated the genetic structure of a large collection of inbred lines and found that 1) heterotic groups are dynamic and evolve over time, 2) small and different sets of inbred lines were critical in contributing to the

parentage of the next era and, 3) a very large proportion of inbred lines did not contribute to the expected genetic structure as a result of selection for hybrid performance. A kernel smoothed Fst scan revealed a small set of loci potentially under directional or balancing selection. Two gene models were implicated that likely coincide with selection response to different biotic and abiotic stresses for bacterial and fungal tolerance and cold tolerance.

4.7 References

Akey, J.M. 2009. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19:711-722.

Anderson, E., and W.L. Brown. 1952. Origin of corn belt maize and its genetic significance. Heterosis: A Record of Researches Directed Toward Explaining and Utilizing the Vigor of Hybrids. Iowa State College Press, Ames, IA. 124-148.

Anon. 1949. Minutes of the meeting of the North Central Corn Breeding Technical Committee. USDA. Beltsville, Maryland.

Anon. 1987. Minutes of the meeting of the North Central Corn Breeding Technical Committee. USDA. Beltsville, Maryland.

Baker, R. 1984. Some of the open pollinated varieties that contributed the most to modern hybrid corn. Illinois Corn Breeder's School Proc. 201-19.

Beissinger, T.M., C.N. Hirsch, B. Vaillancourt, S. Deshpande, K. Barry, C.R. Buell, S.M. Kaeppler, D. Gianola, and N. de Leon. 2014. A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. Genetics 196:829-840.

Black IV, W.C., C.F. Baer, M.F. Antolin, and N.M. DuTeau. 2001. Population genomics: Genome-wide sampling of insect populations. Annu. Rev. Entomol. 46:441-469.

Brown, W.L., and E. Anderson. 1948. The southern dent corns. Ann. Mo. Bot. Gard. 255-268.

Brown, W.L., and E. Anderson. 1947. The northern flint corns. Ann. Mo. Bot. Gard. 1-29.

Comstock, R.E., H. Robinson, and P. Harvey. 1949. A breeding procedure designed to make maximum use of both general and specific combining ability. Agron. J. 41:360-367.

Crow, J.F. 1999. Dominance and Overdominance. p. 49-58. *In* J.G. Coors and S. Pandey (ed) Genetic and exploitation of heterosis in crops. ASA, CSSA, SSSA, Madison, WI.

Duvick, D.N., J. Smith, and M. Cooper. 2004. Long-term selection in a commercial hybrid maize breeding program. Plant Breed. Rev. 24:109-152.

Edwards, J.W., and K.R. Lamkey. 2002. Quantitative genetics of inbreeding in a synthetic maize population. Crop Sci. 42:1094-1104.

Ganal, M.W., G. Durstewitz, A. Polley, A. Bérard, E.S. Buckler, A. Charcosset, J.D. Clarke, E. Graner, M. Hansen, and J. Joets. 2011. A large maize (zea mays L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PloS One 6:e28334.

Gerdes, J.W. and W.F. Tracy. 1993a. Pedigree diversity within the Lancaster Surecrop Heterotic group of maize. Crop. Sci. 33:334-347.

Gerdes, J., W. Tracy, J. Coors, J. Geadlemann, and M.K. Viney. 1993b. Compilation of north american maize breeding germplasm. Crop Science Society of America Madison, WI.

Hallauer, A.R., and J. Miranda. Fo. 1988. Quantitative genetics in maize breeding. Iowa State Univ Press, Ames, IA.

Hartl, D.L., and A.G. Clark. 1997. Principles of population genetics. Sinauer associates Sunderland, MA.

Henderson, C.B. 1976. Maize research and Breeders Manual No. VIII. Illinois Foundation Seeds, Inc. Champaign, Illinois.

Henderson, C.B. 1980. Maize research and Breeders Manual No. XI. Illinois Foundation Seeds, Inc. Champaign, Illinois.

Henderson, C.B. 1984. Maize research and Breeders Manual No. X. Illinois Foundation Seeds, Inc. Champaign, Illinois.

Hermisson, J., and P.S. Pennings. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335-2352.

Hinze, L.L., S. Kresovich, J.D. Nason, and K.R. Lamkey. 2005. Population genetic diversity in a maize reciprocal recurrent selection program. Crop Sci. 45:2435-2442.

Ho, J., S. Kresovich, and K. Lamkey. 2005. Extent and distribution of genetic variation in US maize. Crop Sci. 45:1891-1900.

Hohenlohe, P.A., S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson, and W.A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics 6:e1000862.

Hufford, M.B., X. Xu, J. Van Heerwaarden, T. Pyhäjärvi, J. Chia, R.A. Cartwright, R.J. Elshire, J.C. Glaubitz, K.E. Guill, and S.M. Kaeppler. 2012. Comparative population genomics of maize domestication and improvement. Nat. Genet. 44:808-811.

Johnstone, I., Z. Ma, P. Perry, and M. Shahram. 2009. RMTstat: Distributions, statistics and tests derived from random matrix theory. R Package Version 0.2.

Jones, D.F. 1918. The effect of inbreeding and crossbreeding upon development. Connecticut Agric. Exp. Stn. Bull. 207:5-100.

Labate, J.A., K.R. Lamkey, S.E. Mitchell, S. Kresovich, H. Sullivan, and J.S. Smith. 2003. Molecular and historical aspects of corn belt dent diversity. Crop Sci. 43:80-91.

Lamkey, C., and A. Lorenz. 2014. Relative effect of drift and selection in diverging populations within a reciprocal recurrent selection program. Crop Sci. 54:576-585.

Lauer, J.G., P.R. Carter, T.M. Wood, G. Diezel, D.W. Wiersma, R.E. Rand, and M.J. Mlynarek. 1999. Corn hybrid response to planting date in the northern corn belt. Crop. Sci. 91:834-839.

Liu, K., M. Goodman, S. Muse, J.S. Smith, E. Buckler, and J. Doebley. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117-2128.

Melchinger, A.E., and R.K. Gumber. 1998. Overview of heterosis and heterotic groups in agronomic crops. Concepts and Breeding of Heterosis in Crop Plants. 29-44.

Mikel, M.A., and J.W. Dudley. 2006. Evolution of north american dent corn from public to proprietary germplasm. Crop Sci. 46:1193-1205.

Mock, J.J. and S.A. Eberhart. 1972. Cold tolerance in adapted maize populations. Crop. Sci. 12:466-469

Nelson, P.T., N.D. Coles, J.B. Holland, D.M. Bubeck, S. Smith, and M.M. Goodman. 2008. Molecular characterization of maize inbreds with expired US plant variety protection. Crop Sci. 48:1673-1685.

Patterson, N., A.L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genetics 2:e190.

Pray, L.A., and C.J. Goodnight. 1995. Genetic variation in inbreeding depression in the red flour beetle tribolium castaneum. Evolution 176-188.

Revilla, P., R.A. Malvar, M.E. Cartea, A. Butron, and A. Ordas. 2000. Inheritance of cold tolerance at emergence and during early season growth in maize. Crop Sci. 40:1579-1585.

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, R.J. Elshire, C.B. Acharya, S.E. Mitchell, and S.A. Flint-Garcia. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14:R55.

Shull, G.H. 1908. The composition of a field of maize. J. Hered. 296-301.

Smith, J.S.C., D.N. Duvick, O.S. Smith, M. Cooper, and L. Feng. 2004. Changes in pedigree backgrounds of pioneer brand maize hybrids widely grown from 1930 to 1999. Crop Sci. 44:1935-1946.

Tracy, W., and M. Chandler. 2006. The historical and biological basis of the concept of heterotic patterns in corn belt dent maize. p. 219-233. *In* The historical and biological basis of the concept of heterotic patterns in corn belt dent maize. Plant breeding: The arnel R. hallauer international symposium, 2006. Wiley Online Library, .

Troyer, A.F. 1999. Background of US hybrid corn. Crop Sci. 39:601-626.

van Heerwaarden, J., M.B. Hufford, and J. Ross-Ibarra. 2012. Historical genomics of north american maize. Proc. Natl. Acad. Sci. U. S. A. 109:12420-12425.

Weir, B.S., and C. Cockerham. 1996. Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc. Inc., Sunderland, MA, USA.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 1358-1370.

Wright, S. 1978. Vol. 4: Variability within and among natural populations. University of Chicago Press, Chicago, IL.

4.8 Tables

Table 4.1. Summary of experimental material, time period, significant eigenvalues, genetic groups, and number of individuals per era.

Era	Time period	Decade(s)	Significant eigenvalues	Groups	Number individuals
1	DC	1930 - 1940	2	2	59
2	TWC to SC	1950 - 1960	4	3	146
3	SC	1970	3	5	139
4	ex-pvp	1980	3	5	226

DC = Double cross hybrid
TWC = Three-way cross hybrid
SC = Single cross hybrid
ex-PVP = ex plant variety protection

Table 4.2. Evidence for balancing selection across the four Eras. Gene diversity (D) and Fst are presented for kernel smoothed window across the four eras. Eras, one, two, three, and four correspond to the double-cross, three-way cross to single-cross, single-cross, and advanced single-cross eras, respectively.

Window			D				Fst			
Chr.	Start	End	Era 1	Era 2	Era 3	Era 4	Era 1	Era 2	Era 3	Era 4
1	123,318,407	123,744,703	0.358	0.383	0.373	0.378	0.003	0.030	0.103	0.256
3	84,577,896	84,982,812	0.259	0.292	0.288	0.289	0.010	0.090	0.149	0.070
6	43,537,707	43,909,243	0.263	0.273	0.257	0.273	0.002	0.093	0.087	0.087
6	43,621,222	44,008,995	0.259	0.270	0.256	0.270	0.000	0.102	0.083	0.084
8	174,619,594	174,982,797	0.277	0.295	0.286	0.277	0.011	0.092	0.109	0.081
8	174,760,317	175,150,859	0.269	0.270	0.264	0.257	0.004	0.084	0.111	0.081
10	17,282,613	17,678,034	0.294	0.283	0.284	0.271	0.018	0.032	0.092	0.140
10	73,361,670	73,748,490	0.301	0.297	0.288	0.288	0.027	0.116	0.063	0.133

4.9 Figures

Figure 4.1. Plot of eigenvectors for all inbred lines. Eigenvector one, two and three explained 2.89 %, 2.21 %, and 1.95 % of the variation, respectively. Each era is assigned a different color.

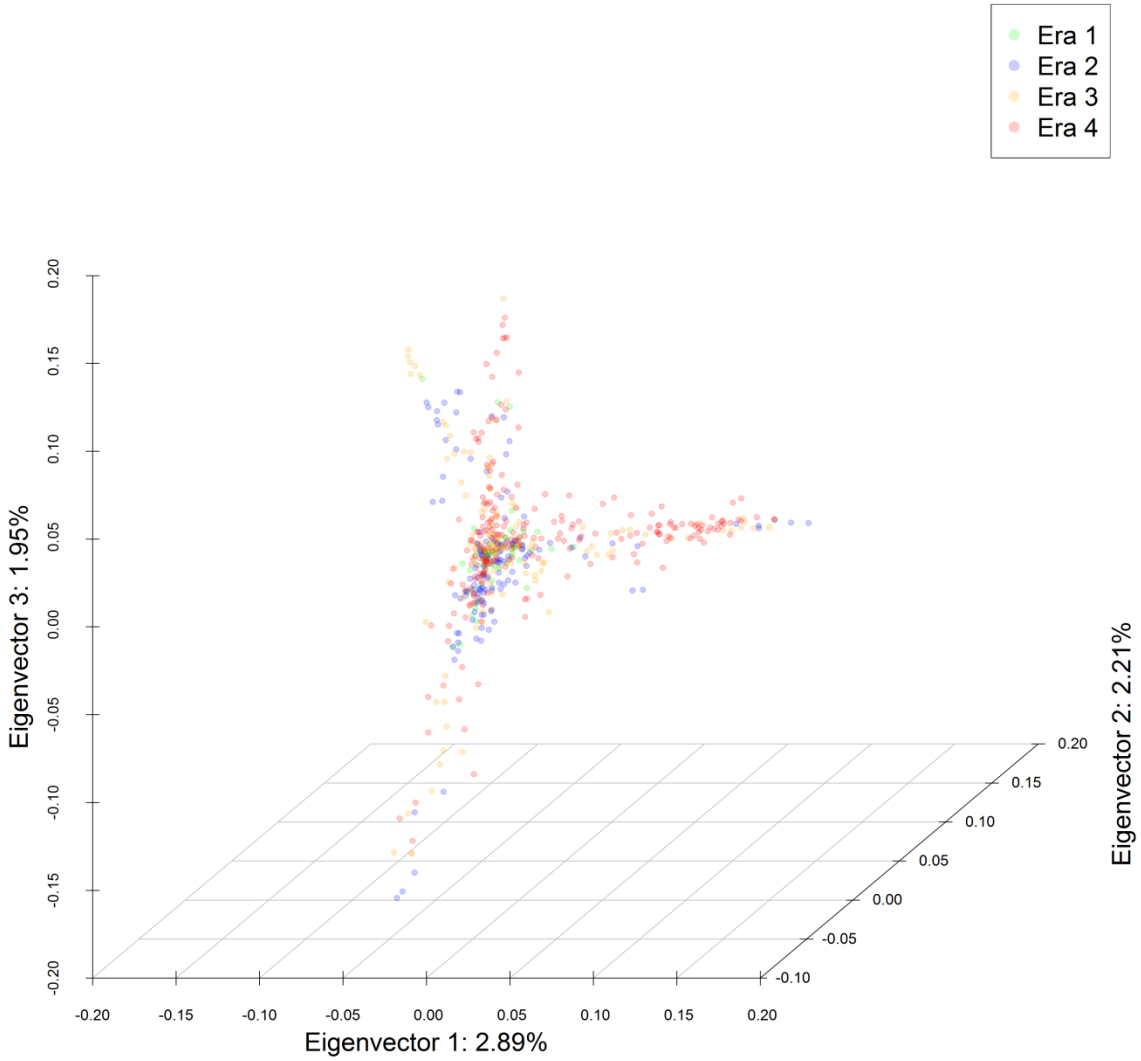


Figure 4.2. Plot of eigenvectors for Era 1 which contained inbred lines developed in 1930 and 1940. Eigenvectors one and two explained 3.42 % and 3.29 % of the variation, respectively.

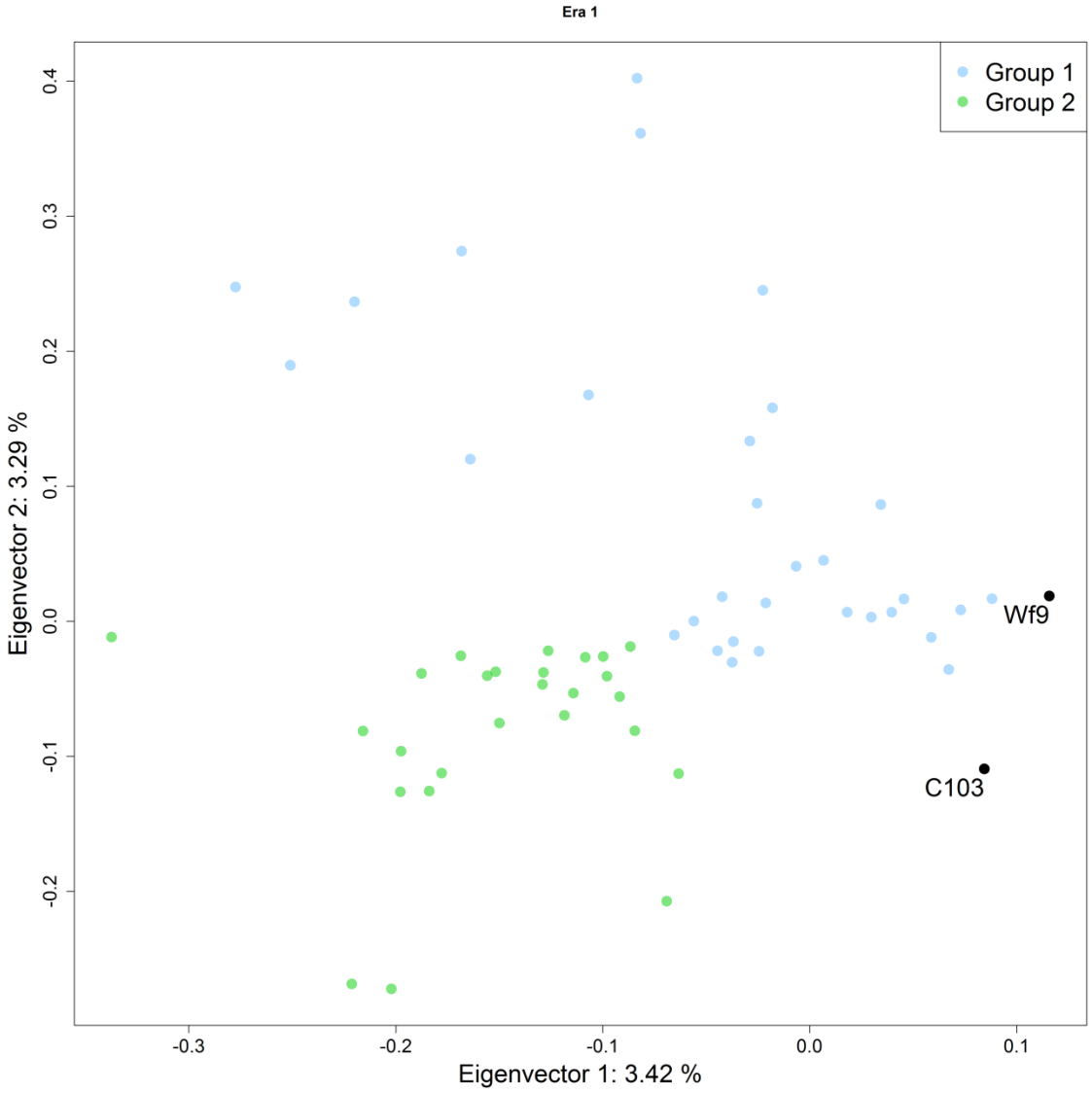


Figure 4.3. Plot of eigenvectors for Era 2 which contained inbred lines developed in 1950 and 1960. Eigenvectors one, two, and three explained 3.7 %, 3.09 %, and 2.69 % of the variation, respectively.

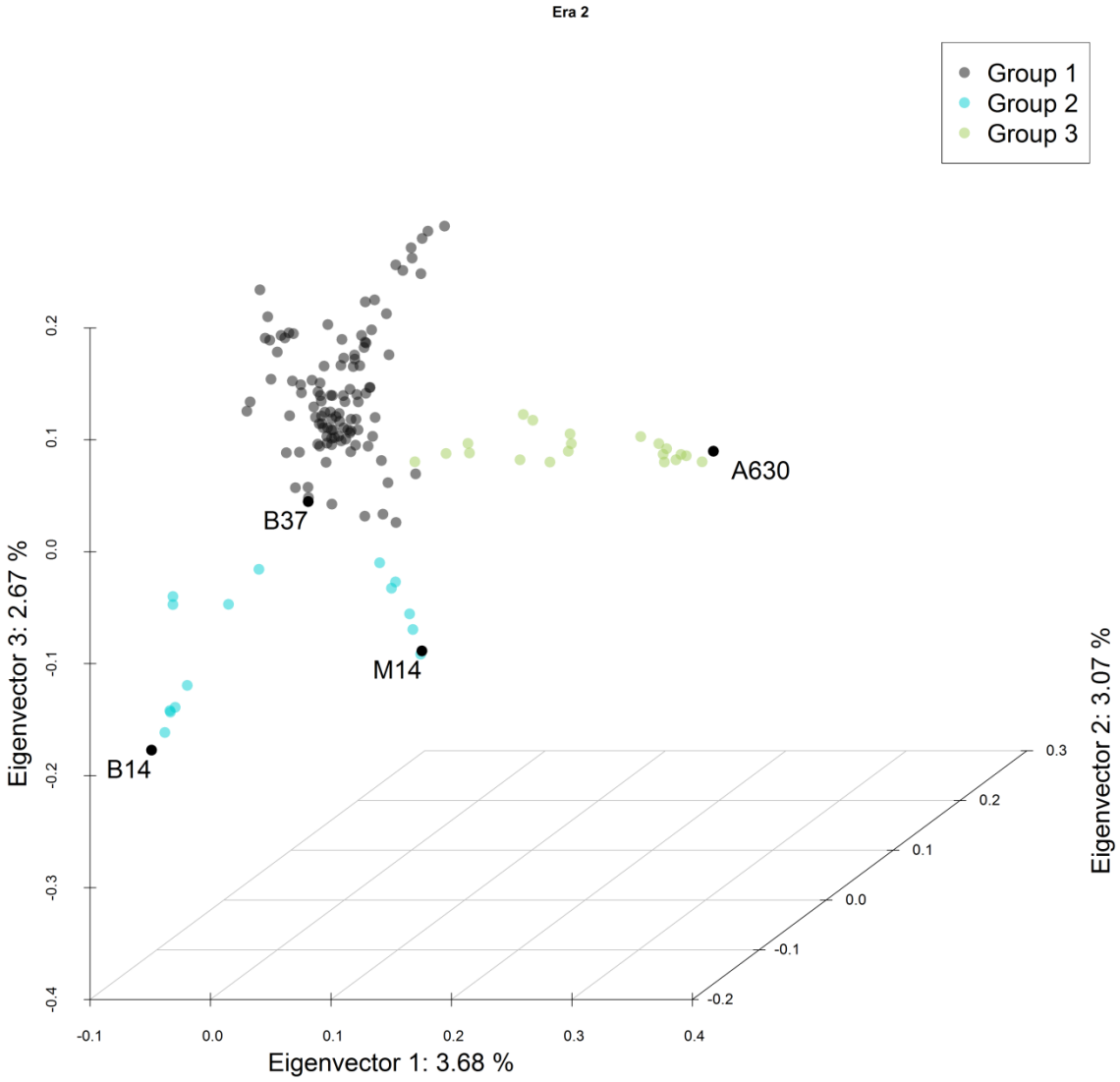


Figure 4.4. Plot of eigenvectors for Era 3 which contained inbred lines developed in 1970. Eigenvectors one, two, and three explained 5.17 %, 3.00 %, and 2.58 % of the variation, respectively.

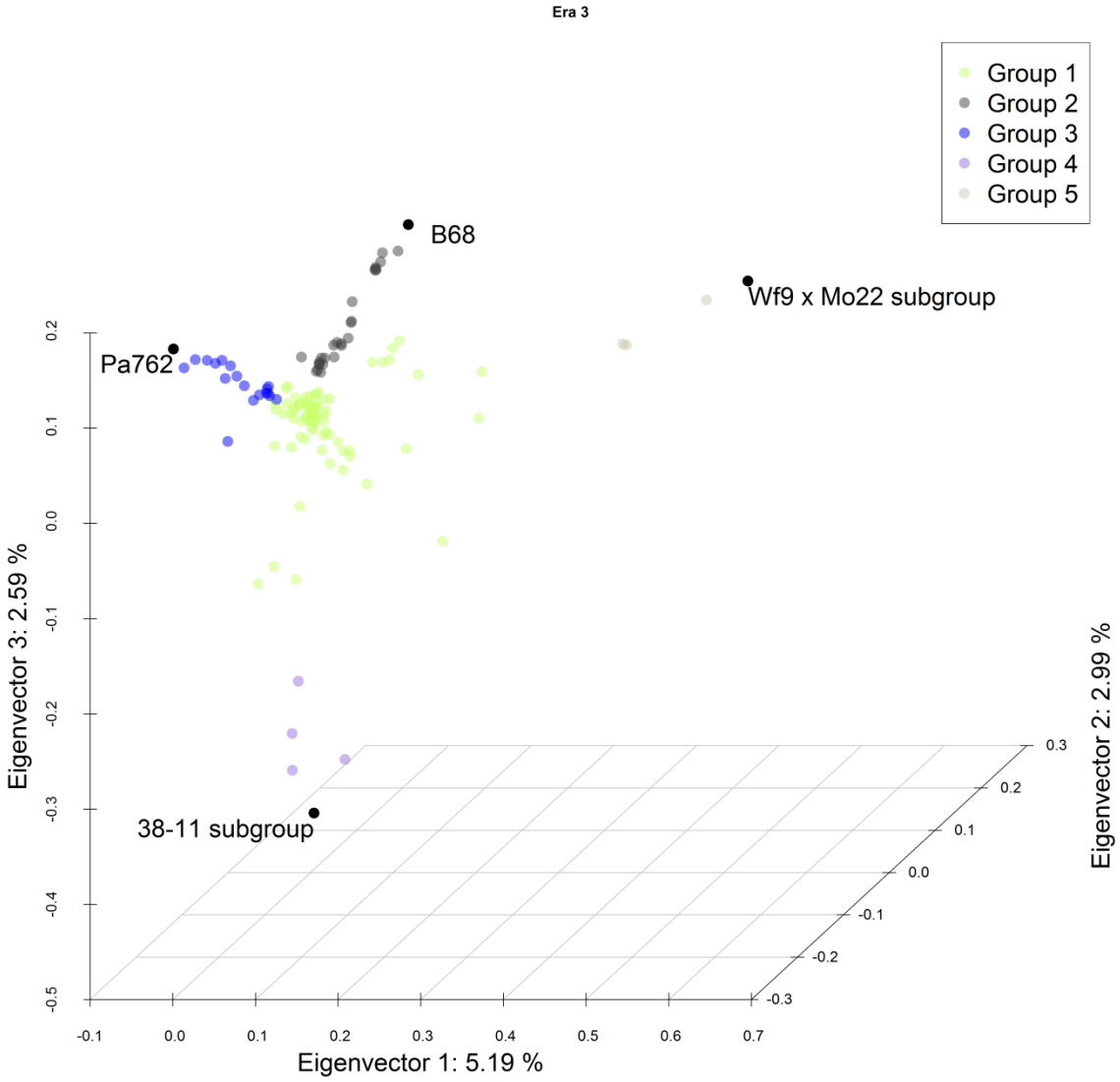


Figure 4.5. Plot of eigenvectors for Era 4 which contained inbred lines developed in 1980. Eigenvectors one, two, and three explained 5.29 %, 4.08 %, and 3.41 % of the variation, respectively.

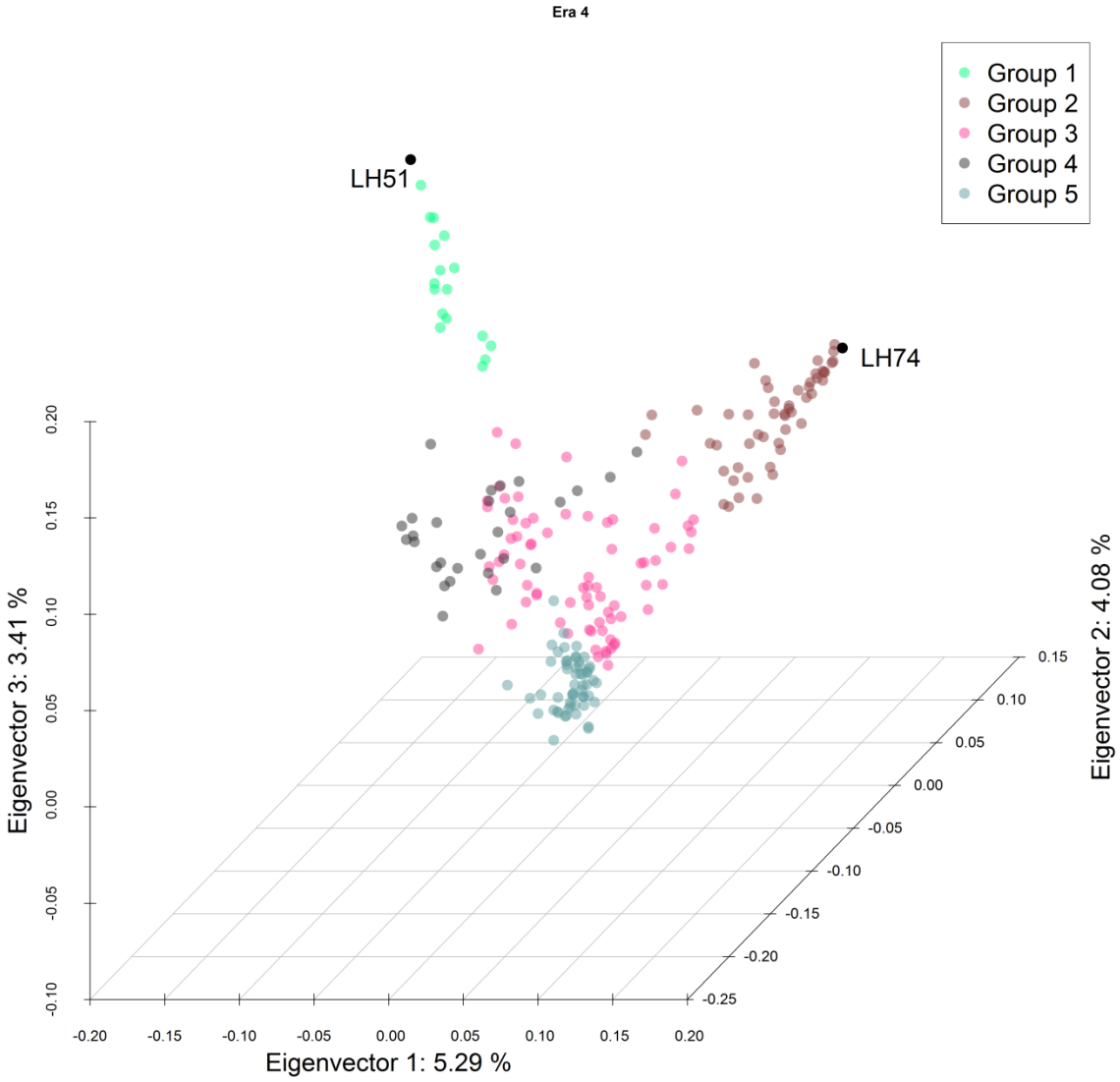


Figure 4.6. A) Density of kernel smoothed Fst values by Era. B) Barplot of the mean kernel smoothed Fst value of each era. Eras were designated one, two, three, and four, and corresponded to inbred lines prevailent in the double-cross, three-way cross to single-cross, single-cross, and advanced single-cross eras of maize breeding, respectively.

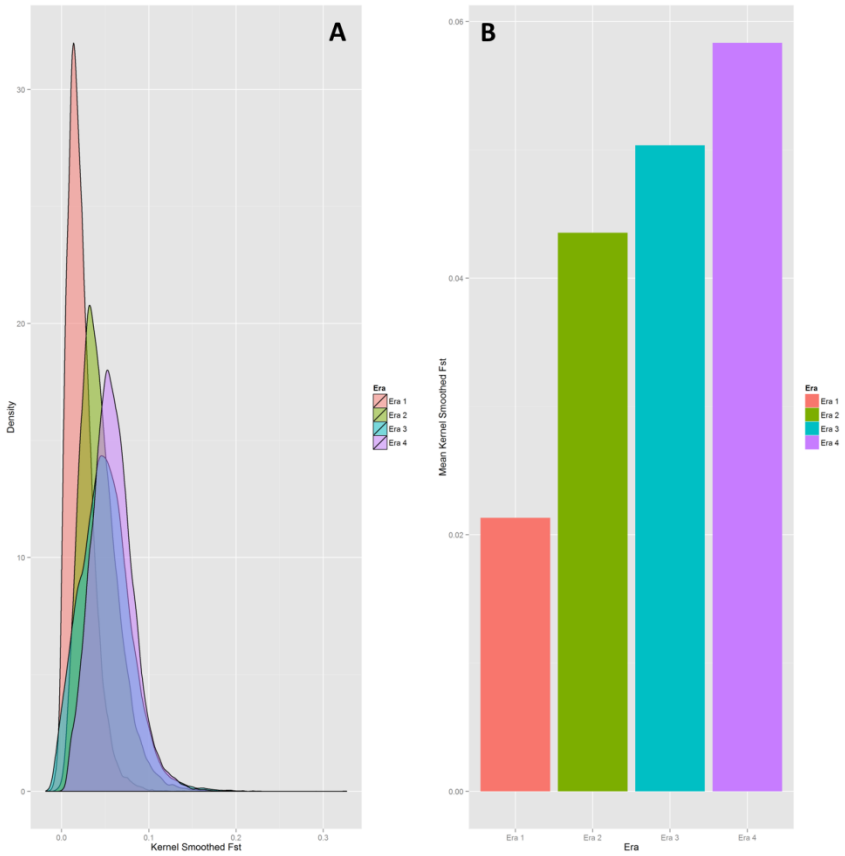


Figure 4.7. Manhattan plot of kernel smoothed (windowsize=400 Kb, stepsize=100Kb) Fst values by Era. A) Era 1, B) Era 2, C) Era 3 and D) Era4. Eras were designated one, two, three, and four, and corresponded to inbred lines prevailent in the double-cross, three-way cross to single-cross, single-cross, and advanced single-cross eras of maize breeding, respectively.

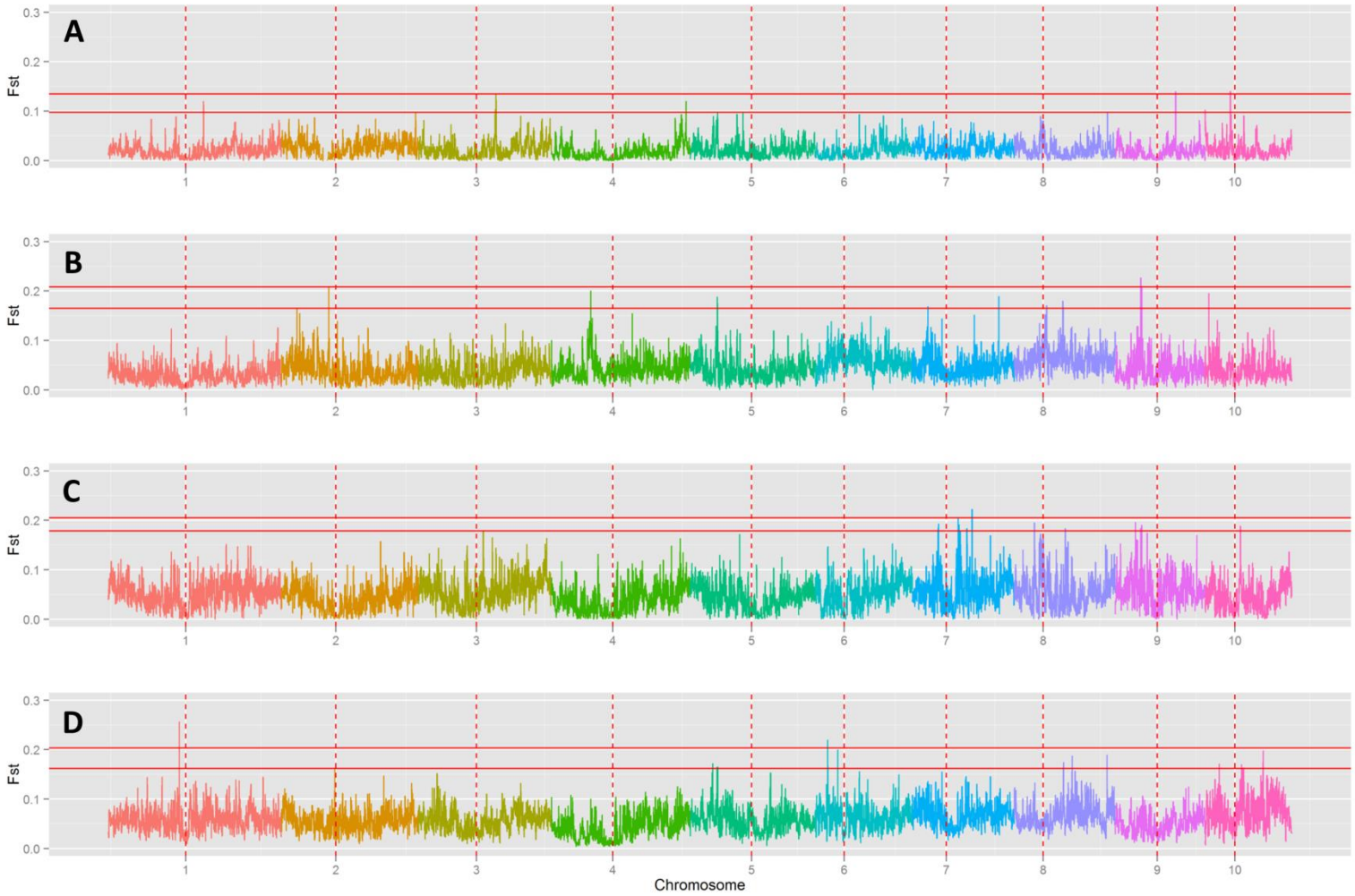
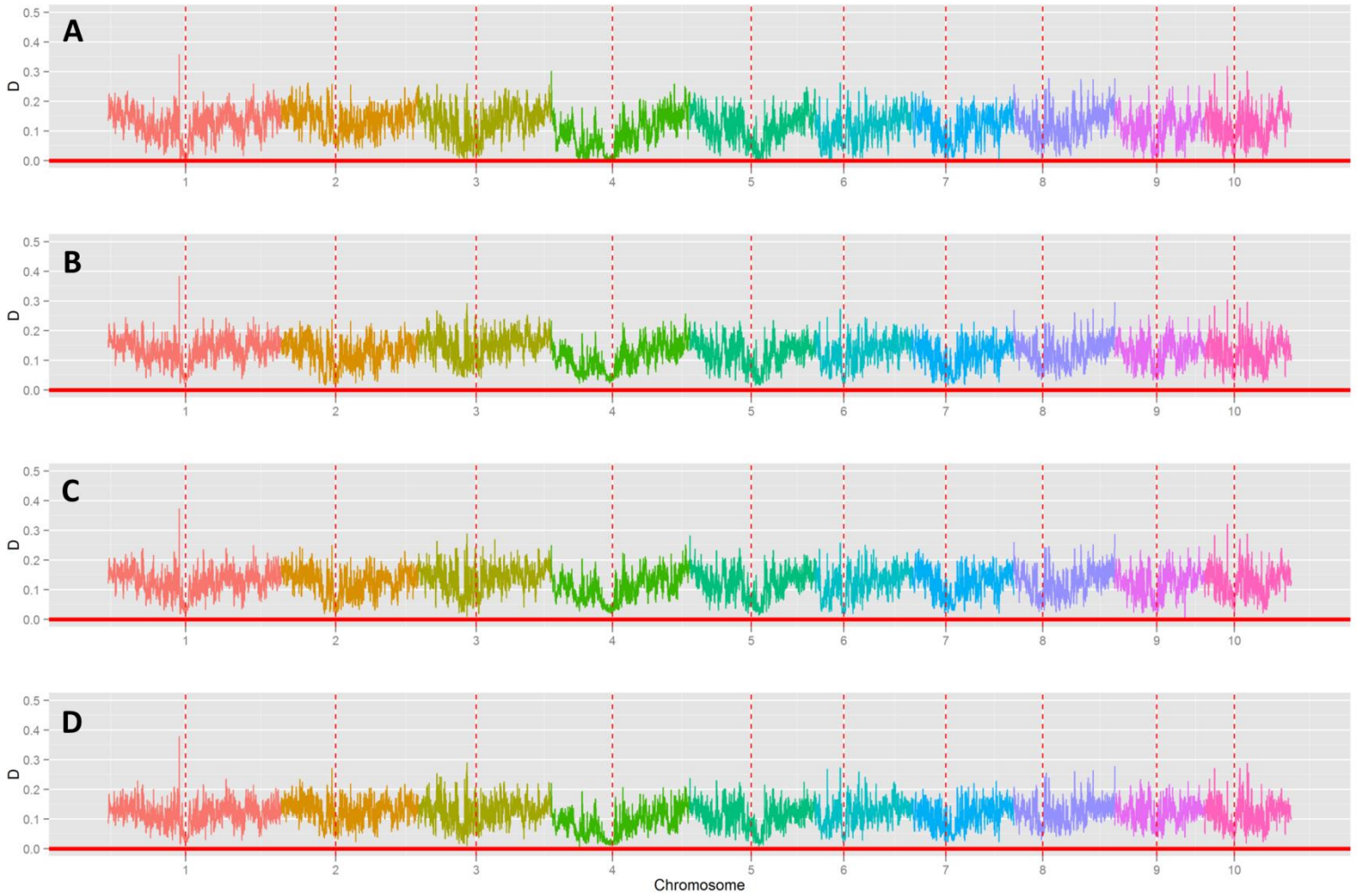

A
B
C
D
Fst
0.0
0.1
0.2
0.3
1
2
3
4
5
6
7
8
9
10
Chromosome

Figure 4.8. Manhattan plot of kernel smoothed (windowsize=400 Kb, stepsize=100Kb) gene diversity (D) values by Era. A) Era 1, B) Era 2, C) Era 3 and D) Era4. Eras were designated one, two, three, and four, and corresponded to inbred lines prevalent in the double-cross, three-way cross to single-cross, single-cross, and advanced single-cross eras of maize breeding, respectively.


A
B
C
D
D
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
5
6
7
8
9
10
Chromosome

5 CHAPTER 5 CONCLUSIONS

This dissertation contributes to the currently small body of literature investigating the relative impacts of selection and drift in the formation of population structure. We attempt to characterize the impacts of selection and genetic drift genome-wide in chapter two and across the genome in chapter three. In chapter four, we switch gears and characterize the formation of population structure in a large collection of inbred lines and search for evidence of selection. Collectively, these results strengthen notions already put forth about the formation of heterotic groups and the relative contributions of different germplasm to heterotic groups and corroborate the research of others regarding the impacts of selection and genetic drift across the genome.

We found no evidence to suggest that selection for hybrid performance has left a distinctive signature of selection genome-wide or at specific locations. At first this may seem surprising, however, it is likely many of the sites under selection conform with a soft-sweep model of adaptation or even a polygenetic model of adaptation. In each case a bulk of the sites under selection will go undetected. I also suggest that multi-allelic systems introduce complications into accurately representing the genotype in the population cross. This issue is more difficult to provide evidence for. Current genotype calling methodology coupled with current sequencing technology make calling any more the two alleles per site difficult because the probability of accurately calling third or even a fourth allele at a site is low due to sequencing errors. Confounding this issue is the

heterozygous nature of the individuals constituting the populations adding to the difficulty in accurately calling more than two alleles. If multi-allelic systems are at play, selection will be largely ineffective in the population cross until genetic drift reduces diversity to the point that a stable genotype appears.

In the context of genome-wide selection scans the built-in replication of the UNL-RpRS program allows for validation of putative selection candidates. At the same time, when selection consistently impacts similar genomic regions across replicates this provides evidence for parallel evolution. Parallel evolution conforms to a soft-sweep model of evolution implying the source of variation for adaptation originates from standing genetic variation. Soft-sweep models of evolution can leave a very slight selection signature behind and are difficult to detect. This evidence coupled with evidence from the literature suggests a soft-sweep model of adaptation is common in maize breeding. A search for evidence of directional selection revealed 10 regions that were under selection in at least two different replicates. A literature review revealed that one of these regions located on chromosome seven is putatively associated with selection for hybrid performance and heterosis.

In the final chapter we characterized a large set of inbred lines split into four eras. Consistent with existing body of literature characterizing heterotic groups in maize, we found similar sets of inbred lines, which have been historically important in contributing to the parentage for the next cycle of selection. A scan for putative selection candidates revealed regions under directional selection and a different set of regions under a mode of

selection maintaining diversity. Two regions were associated with genes putatively controlling cold tolerance and tolerance to bacterial and fungal pathogens.

6 APPENDIX A

In order to curate the supplementary information associated with each chapter, the appendix has been divided into three subsections 6.1, 6.2, and 6.3, corresponding to chapters two, three, and four, respectively.

6.1 Chapter 2 Supplementary Information

6.1.1 Supplementary Table 1

Supplementary Table 1. List of progenitor lines of Nebraska B Synthetic (NBS) and Nebraska Stiff Stalk Synthetic (NSS). References for this list are Gerdes et al (1993), Hagdorn et al., (2003), Odhiambo (1987), and Troyer (2004).

Population	Line	Pedigree/Background	State
NBS	CC16	Northwestern Dent	Wisconsin
NBS	CC7	Golden Glow	Wisconsin
NBS	I234	Iodent Reid	Iowa
NBS	A48	Funk Yellow Dent	Illinois
NBS	D6	Unknown	Illinois
NBS	Oh2614	Hogue Yellow Dent Lines	Ohio
NBS	Oh5414	Unknown	Ohio
NBS	Ohio O4	Unknown	Ohio
NBS	US4-8	Lancaster Surecrop	Iowa
NBS	CI540	Il Two Ear	USDA
NBS	Oh28	Funk Yellow Dent	Ohio
NBS	Oh608B	Hogue Yellow Dent	Ohio
NBS	K	Hayes Golden	Illinois
NBS	90	Funk 90 day	Illinois
NBS	ITE701	Illinois Two Ear	Iowa
NBS	5120	Illinois High Yield	Illinois
NBS	530	Unknown	Illinois
NBS	Pr	Proudfit Reid	Iowa
NBS	Os426	Osterland Yellow Dent	Iowa
NBS	A374	Reid Yellow Dent	Minnesota
NBS	A375	Reid Yellow Dent	Minnesota
NBS	Oh51A	Wooster Clarage	Ohio
NBS	I198	Iodent Reid	Iowa
NBS	P8	Palin Reid	Indiana
NBS	Tr	Troyer Reid	Indiana
NBS	R4	Funk Yellow Dent	Illinois
NBS	Oh40B	Lancaster	Ohio
NBS	K4	Kansas Sunflower	Kansas
NBS	CC5	Golden Glow	Wisconsin
NBS	A	Funk Yellow Dent	Illinois
NBS	L	Mann Leaming	Illinois
NBS	M14	Funk Yellow Dent x TexasSurecropper	Illinois
NSS	A3G-3-1-3	BL345B x I159	
NSS	CI617	Funk Reid	Iowa
NSS	F1B1-7-1	Troyer Reid	
NSS	LE23	Illinois Low Ear, ChesterLeaming	Illinois
NSS	461-3	Troyer Reid	Indiana
NSS	I159	Iodent Reid	Iowa
NSS	CI187-2	Krug Reid x Gold Mine	USDA
NSS	Hy	Illinois High Yield	Illinois

NSS	I224	Iodent Reid	Iowa
NSS	WD456	Walden Reid	Iowa
NSS	12E	Kansas Sunflower	Illinois
NSS	AH83	Funk Reid	Indiana
NSS	Tr9-1-1-6	Troyer Reid	Indiana
NSS	Oh3167B	Echelberger Clarage	Ohio
NSS	CI540	Illinois Two-ear	USDA
NSS	Os420	Osterland Yellow Dent	Iowa

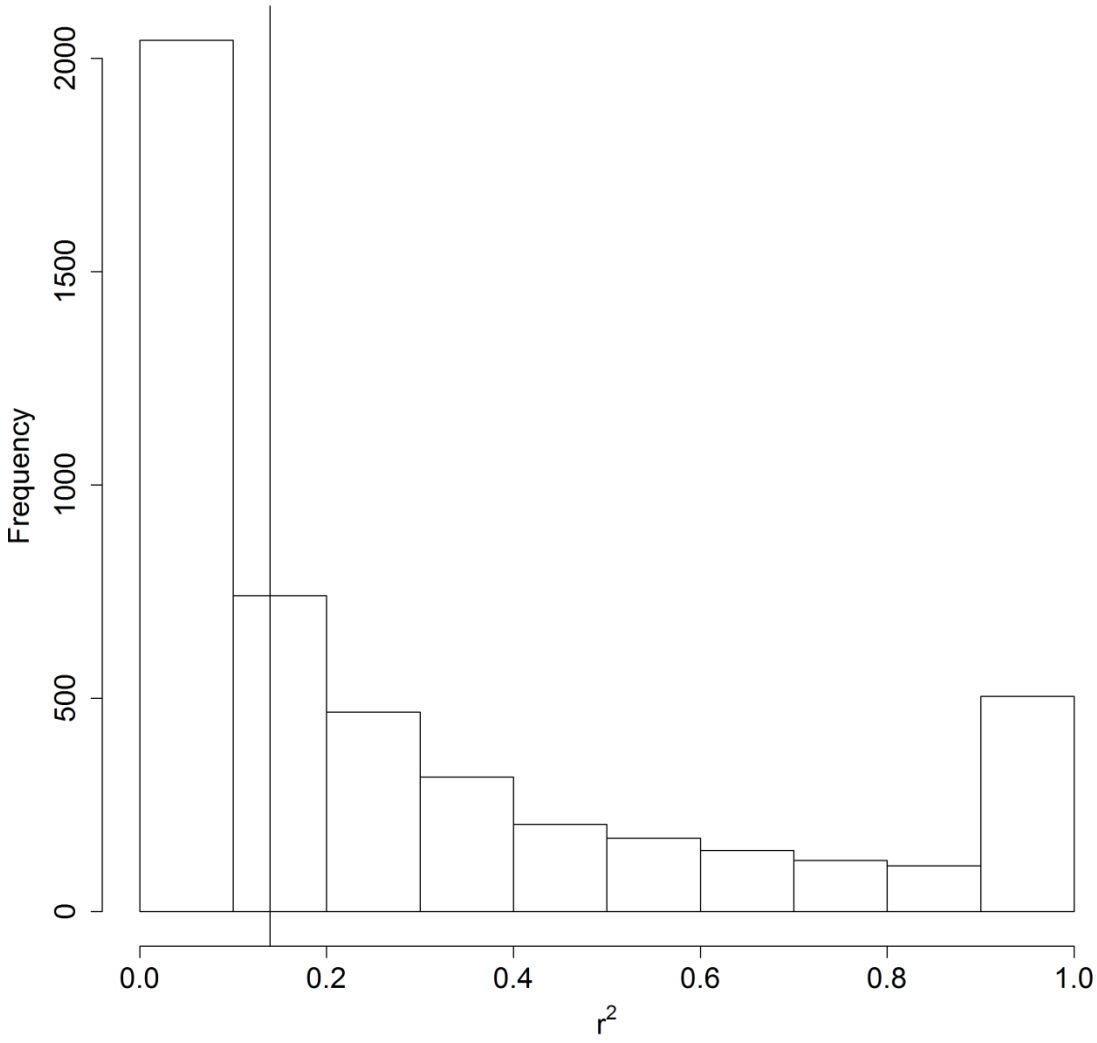
6.1.2 Supplementary Table 2

Supplementary Table 2. Number of polymorphic markers in linkage disequilibrium ($r^2 > 0.20$) with at least one adjacent locus. Observed heterozygosity (Ho) and gene diversity (D) for the 14 UNL-RpRS populations are displayed.

Population	Polymorphic markers	Ho	D
NBS_C0	192	0.304	0.298
NBS_RFS_C8_1	171	0.225	0.233
NBS_RFS_C8_2	117	0.198	0.184
NBS_RFS_C8_3	140	0.221	0.215
NBS_S1_C8_1	136	0.226	0.209
NBS_S1_C8_2	118	0.183	0.173
NBS_S1_C8_3	98	0.156	0.151
NSS_C0	217	0.354	0.339
NSS_RFS_C8_1	152	0.235	0.229
NSS_RFS_C8_2	150	0.229	0.221
NSS_RFS_C8_3	140	0.223	0.204
NSS_S1_C8_1	163	0.2482	0.231
NSS_S1_C8_2	192	0.228	0.251
NSS_S1_C8_3	157	0.226	0.230
Std. Err	--	0.005	0.007

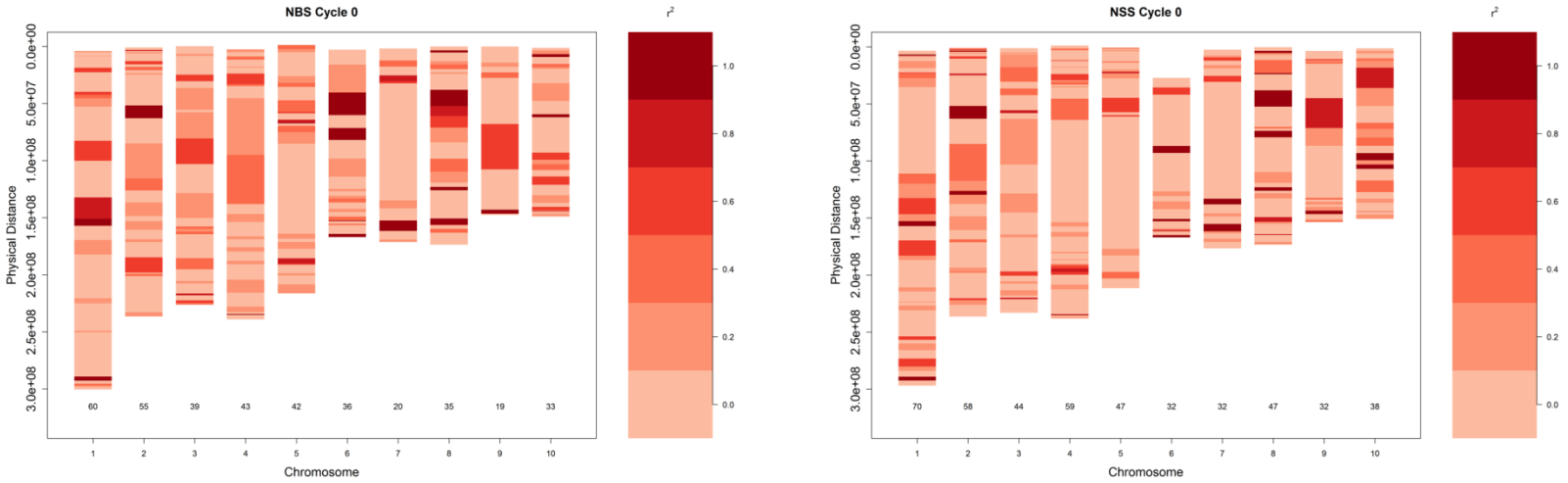
6.1.3 Supplementary Figure 1

Histogram of adjacent linkage disequilibrium for 14 populations in the University of Nebraska replicated recurrent selection program. The vertical line is the mean adjacent linkage disequilibrium (0.29).



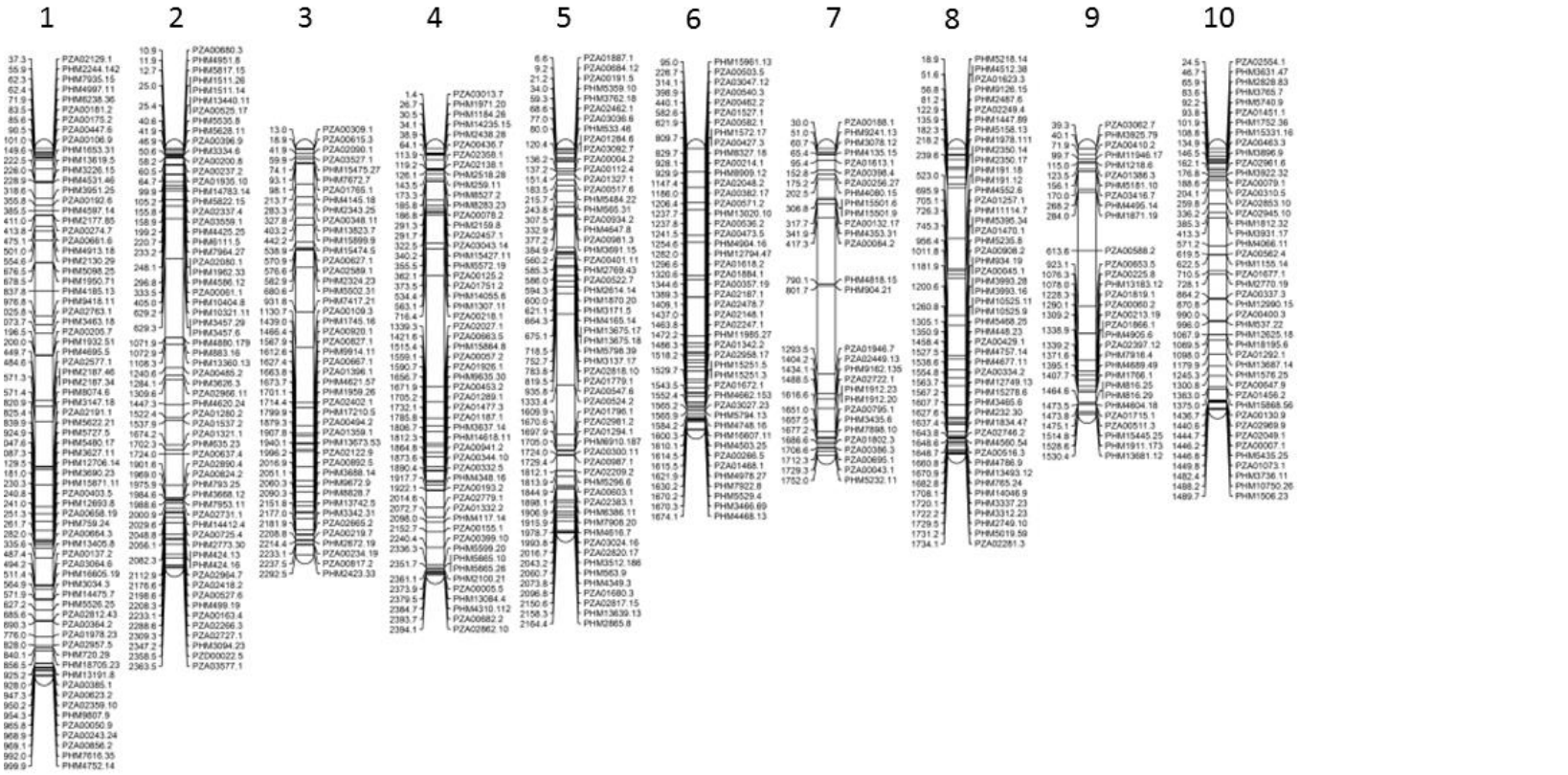
6.1.4 Supplementary Figure 2

Visualization physical distance verse chromosome for adjacent linkage disequilibrium for Nebraska B Synthetic (NBS) and Nebraska Stiff Stalk Synthetic (NSS) cycle 0 populations. The value along the x-axis indicates the number of markers per chromosome.



6.1.5 Supplementary Figure 3

Physical map of the 513 markers used for analysis. Physical distance were divided by 100 000 due to space constraints. Multiply by 100 000 for physical position.



6.2 Chapter 3 Supplementary Information

6.2.1 Supplementary Table 1.

List of selection candidates identified by an outlier test at the 99.9 % level for C0 and C8 population comparisons. This table includes the population comparison, chromosome (CHR), window start (Window_Start), window stop (Window_End), number of SNPs per window (SNPs), and Fst value. Population comparisons are in abbreviated format and interpetted as follows: the first two letters denote base population NB (Nebraska B Synthetic) or NS (Nebraska Stiff Stalk), the next letter and number indicate cycle 0 (C0), The remaining spots indicate the cycle 8 population. For example NBC0RFS1 indicates Fst was computed between NBS_C0 population and NBS_RFS_C8_1 population.

Comparison	CHR	Window_Start	Window_End	SNPs	Fst
NBC0RFS1	1	86,730,202	87,169,988	11	0.246
NBC0RFS1	1	91,267,367	91,768,087	60	0.241
NBC0RFS1	4	128,263,401	128,603,647	19	0.274
NBC0RFS1	4	128,263,415	128,860,445	18	0.285
NBC0RFS1	4	128,477,152	128,874,332	21	0.240
NBC0RFS1	4	131,938,210	132,213,073	4	0.352
NBC0RFS1	5	26,444,405	26,774,088	9	0.333
NBC0RFS1	6	71,449,535	72,008,063	23	0.311
NBC0RFS1	7	17,693,465	18,147,035	8	0.245
NBC0RFS1	8	80,791,959	81,247,853	22	0.236
NBC0RFS1	8	81,160,136	81,639,109	24	0.260
NBC0RFS1	8	109,378,843	109,910,362	17	0.262
NBC0RFS1	9	45,850,091	46,385,864	4	0.238
NBC0RFS1	9	64,411,571	64,808,356	18	0.265
NBC0RFS1	9	64,572,545	65,200,562	18	0.268
NBC0RFS1	9	65,962,331	66,686,105	2	0.243
NBC0RFS2	1	61,719,602	62,142,811	44	0.264
NBC0RFS2	3	63,253,384	63,696,416	10	0.244
NBC0RFS2	3	73,639,262	74,096,740	7	0.326
NBC0RFS2	3	73,639,309	74,371,255	7	0.321
NBC0RFS2	3	73,784,151	74,096,740	5	0.452
NBC0RFS2	3	129,565,899	130,049,835	22	0.233
NBC0RFS2	8	51,064,781	51,485,491	14	0.235
NBC0RFS2	8	58,026,451	58,532,024	2	0.314

NBC0RFS2	8	61,031,438	61,521,622	11	0.311
NBC0RFS2	10	19,578,402	19,853,527	11	0.259
NBC0RFS2	10	58,673,233	59,167,459	9	0.302
NBC0RFS2	10	59,003,246	59,448,899	12	0.272
NBC0RFS2	10	64,268,829	64,860,438	9	0.508
NBC0RFS2	10	64,598,799	64,993,245	29	0.357
NBC0RFS2	10	64,598,834	65,099,531	24	0.290
NBC0RFS2	10	64,860,438	65,340,654	31	0.245
NBC0RFS3	1	176,170,056	176,757,342	2	0.295
NBC0RFS3	2	221,372,354	221,894,582	21	0.257
NBC0RFS3	3	10,626,611	11,060,904	99	0.236
NBC0RFS3	3	127,969,644	128,533,531	26	0.241
NBC0RFS3	3	128,093,045	128,533,550	20	0.241
NBC0RFS3	3	128,099,185	128,638,123	11	0.255
NBC0RFS3	4	126,880,078	127,419,843	19	0.275
NBC0RFS3	5	154,621,730	155,106,707	53	0.233
NBC0RFS3	5	154,621,772	155,244,763	51	0.242
NBC0RFS3	7	113,535,012	114,178,113	3	0.289
NBC0RFS3	7	113,706,281	114,196,433	6	0.369
NBC0RFS3	8	153,561,499	154,054,086	60	0.232
NBC0RFS3	8	153,704,917	154,155,922	83	0.301
NBC0RFS3	8	153,708,856	154,246,271	73	0.326
NBC0RFS3	8	153,858,326	154,303,632	82	0.289
NBC0RFS3	8	153,897,929	154,388,752	87	0.261
NBC0RFS	1	78,364,771	79,005,929	2	0.244
NBC0RFS	1	176,170,056	176,757,342	2	0.246
NBC0RFS	1	242,081,512	242,581,078	34	0.145
NBC0RFS	1	242,161,961	242,708,615	32	0.153
NBC0RFS	2	80,587,113	80,897,091	3	0.205
NBC0RFS	3	162,297,196	162,703,519	47	0.145
NBC0RFS	3	168,301,607	168,767,027	35	0.150
NBC0RFS	3	168,368,508	168,863,064	43	0.178
NBC0RFS	3	168,492,695	168,962,432	53	0.192
NBC0RFS	3	168,525,999	169,030,137	48	0.201
NBC0RFS	4	141,096,518	141,602,728	40	0.191
NBC0RFS	5	135,845,492	136,415,766	14	0.175
NBC0RFS	5	194,839,615	195,419,422	29	0.183
NBC0RFS	5	194,976,927	195,455,621	36	0.144
NBC0RFS	5	194,976,992	195,482,629	35	0.149

NBC0RFS	6	138,987,310	139,454,499	31	0.190
NBC0S11	1	104,276,881	104,679,457	12	0.325
NBC0S11	1	104,276,904	104,813,298	11	0.356
NBC0S11	1	104,608,484	105,024,450	16	0.304
NBC0S11	1	168,432,885	169,005,627	7	0.343
NBC0S11	1	168,693,295	169,356,618	7	0.311
NBC0S11	1	176,170,056	176,757,342	2	0.296
NBC0S11	3	68,358,195	68,752,219	17	0.317
NBC0S11	3	68,359,752	68,905,707	14	0.377
NBC0S11	3	68,752,204	69,376,043	14	0.377
NBC0S11	5	141,258,839	141,580,890	2	0.313
NBC0S11	5	143,198,267	143,672,074	21	0.276
NBC0S11	5	143,400,381	143,958,042	19	0.263
NBC0S11	5	153,946,756	154,481,565	6	0.312
NBC0S11	7	1,318,319	1,794,525	51	0.269
NBC0S11	7	1,441,797	1,938,526	50	0.261
NBC0S11	7	111,733,857	112,219,188	4	0.302
NBC0S12	1	176,170,056	176,757,342	2	0.289
NBC0S12	3	85,717,848	86,197,942	26	0.283
NBC0S12	4	128,263,401	128,603,647	19	0.279
NBC0S12	4	128,263,415	128,860,445	18	0.291
NBC0S12	4	128,477,152	128,874,332	21	0.246
NBC0S12	4	131,938,210	132,213,073	4	0.358
NBC0S12	5	22,103,166	22,603,061	48	0.290
NBC0S12	5	22,150,253	22,637,453	57	0.270
NBC0S12	5	132,657,090	133,189,938	2	0.290
NBC0S12	5	194,839,615	195,419,422	29	0.314
NBC0S12	7	113,535,012	114,178,113	3	0.511
NBC0S12	7	113,706,281	114,196,433	6	0.471
NBC0S12	8	99,419,376	99,994,152	9	0.306
NBC0S12	8	99,590,156	99,994,159	9	0.292
NBC0S12	8	109,378,843	109,910,362	17	0.244
NBC0S12	10	12,754,461	13,334,822	15	0.255
NBC0S13	1	78,364,771	79,005,929	2	0.276
NBC0S13	2	146,706,388	147,035,493	6	0.320
NBC0S13	4	37,390,536	38,022,503	23	0.369
NBC0S13	4	50,100,951	50,550,926	45	0.273
NBC0S13	4	69,921,252	70,354,606	7	0.390
NBC0S13	4	140,349,938	140,878,179	4	0.587

NBC0S13	6	4,894,709	5,441,885	25	0.321
NBC0S13	7	90,183,249	90,519,448	10	0.427
NBC0S13	7	90,246,776	90,789,748	9	0.475
NBC0S13	7	90,519,391	90,789,792	10	0.395
NBC0S13	7	113,535,012	114,178,113	3	0.575
NBC0S13	7	113,706,281	114,196,433	6	0.532
NBC0S13	10	25,668,624	26,135,208	12	0.290
NBC0S13	10	73,205,989	73,748,093	25	0.315
NBC0S13	10	73,283,433	73,748,455	16	0.361
NBC0S13	10	75,814,884	76,406,452	13	0.282
NBC0S1	1	78,364,771	79,005,929	2	0.238
NBC0S1	1	176,170,056	176,757,342	2	0.261
NBC0S1	2	80,587,113	80,897,091	3	0.218
NBC0S1	3	168,301,607	168,767,027	35	0.191
NBC0S1	3	168,492,695	168,962,432	53	0.199
NBC0S1	3	168,525,999	169,030,137	48	0.206
NBC0S1	4	24,422,540	24,950,473	3	0.194
NBC0S1	5	77,527,266	77,998,215	64	0.227
NBC0S1	5	77,658,040	78,255,195	64	0.227
NBC0S1	5	77,760,331	78,255,213	21	0.228
NBC0S1	5	194,839,615	195,419,422	29	0.223
NBC0S1	6	138,987,310	139,454,499	31	0.195
NBC0S1	7	113,535,012	114,178,113	3	0.229
NBC0S1	7	113,706,281	114,196,433	6	0.252
NBC0S1	9	64,572,545	65,200,562	18	0.189
NBC0S1	9	65,962,331	66,686,105	2	0.226
NSC0RFS1	1	138,914,447	139,448,347	6	0.291
NSC0RFS1	2	86,916,123	87,656,448	10	0.314
NSC0RFS1	2	102,285,607	102,753,374	4	0.334
NSC0RFS1	2	205,036,265	205,493,595	10	0.308
NSC0RFS1	3	39,478,477	40,062,732	5	0.526
NSC0RFS1	3	73,639,262	74,096,740	7	0.333
NSC0RFS1	3	73,639,309	74,371,255	7	0.330
NSC0RFS1	3	73,784,151	74,096,740	5	0.443
NSC0RFS1	3	74,441,922	75,095,053	2	0.343
NSC0RFS1	3	122,518,947	122,911,119	5	0.375
NSC0RFS1	4	135,481,568	136,002,833	16	0.303
NSC0RFS1	5	132,044,647	132,497,713	8	0.335
NSC0RFS1	6	30,257,596	30,516,029	3	0.415

NSC0RFS1	6	30,515,985	31,006,173	13	0.370
NSC0RFS1	6	30,516,029	31,035,375	13	0.350
NSC0RFS1	9	51,413,632	51,951,860	4	0.277
NSC0RFS2	2	86,252,304	86,768,599	9	0.305
NSC0RFS2	2	86,793,596	87,277,651	13	0.294
NSC0RFS2	2	86,916,123	87,656,448	10	0.429
NSC0RFS2	2	87,277,538	87,765,514	14	0.321
NSC0RFS2	2	212,644,110	213,100,483	92	0.256
NSC0RFS2	3	73,784,151	74,096,740	5	0.238
NSC0RFS2	3	94,139,430	94,795,670	9	0.231
NSC0RFS2	3	118,495,616	118,968,735	8	0.373
NSC0RFS2	4	18,936,748	19,430,220	22	0.244
NSC0RFS2	4	36,418,456	36,914,505	66	0.241
NSC0RFS2	5	93,593,712	93,897,179	2	0.259
NSC0RFS2	5	179,278,157	179,790,926	7	0.252
NSC0RFS2	6	30,257,596	30,516,029	3	0.271
NSC0RFS2	7	90,183,249	90,519,448	10	0.279
NSC0RFS2	7	90,246,776	90,789,748	9	0.324
NSC0RFS2	8	123,657,634	124,208,973	43	0.266
NSC0RFS3	6	101,015,413	101,513,634	16	0.284
NSC0RFS3	6	101,192,502	101,801,332	16	0.274
NSC0RFS3	6	101,192,553	101,513,634	10	0.340
NSC0RFS3	7	26,599,643	27,032,384	20	0.347
NSC0RFS3	7	27,422,188	27,982,066	8	0.290
NSC0RFS3	7	27,543,280	27,982,091	6	0.293
NSC0RFS3	7	50,591,882	51,101,564	3	0.489
NSC0RFS3	7	90,246,776	90,789,748	9	0.292
NSC0RFS3	7	90,519,391	90,789,792	10	0.256
NSC0RFS3	7	111,733,857	112,219,188	4	0.282
NSC0RFS3	7	113,535,012	114,178,113	3	0.336
NSC0RFS3	7	113,706,281	114,196,433	6	0.288
NSC0RFS3	9	64,411,571	64,808,356	18	0.333
NSC0RFS3	9	64,572,545	65,200,562	18	0.333
NSC0RFS3	9	93,373,030	93,814,213	18	0.250
NSC0RFS3	10	66,957,392	67,227,691	4	0.266
NSC0RFS	1	78,364,771	79,005,929	2	0.213
NSC0RFS	2	86,916,123	87,656,448	10	0.204
NSC0RFS	2	100,258,630	100,584,677	2	0.173
NSC0RFS	2	213,310,725	213,808,630	86	0.176

NSC0RFS	5	194,839,615	195,419,422	29	0.178
NSC0RFS	6	101,015,413	101,513,634	16	0.160
NSC0RFS	6	101,192,502	101,801,332	16	0.165
NSC0RFS	6	101,192,553	101,513,634	10	0.242
NSC0RFS	6	120,279,997	120,692,710	26	0.174
NSC0RFS	6	137,891,561	138,355,859	32	0.155
NSC0RFS	7	20,620,150	21,124,103	31	0.154
NSC0RFS	7	90,183,249	90,519,448	10	0.187
NSC0RFS	7	90,246,776	90,789,748	9	0.225
NSC0RFS	7	90,519,391	90,789,792	10	0.196
NSC0RFS	7	102,534,078	102,869,411	4	0.159
NSC0RFS	7	136,391,056	136,942,543	27	0.174
NSC0S11	2	76,165,257	76,527,130	7	0.311
NSC0S11	2	81,818,927	82,392,420	5	0.423
NSC0S11	2	86,252,304	86,768,599	9	0.329
NSC0S11	2	102,285,607	102,753,374	4	0.313
NSC0S11	3	73,784,151	74,096,740	5	0.357
NSC0S11	7	26,599,643	27,032,384	20	0.329
NSC0S11	7	33,915,623	34,470,637	18	0.392
NSC0S11	7	113,535,012	114,178,113	3	0.361
NSC0S11	7	113,706,281	114,196,433	6	0.321
NSC0S11	9	64,411,571	64,808,356	18	0.297
NSC0S11	9	64,572,545	65,200,562	18	0.297
NSC0S11	10	12,754,461	13,334,822	15	0.332
NSC0S11	10	33,349,113	33,750,298	20	0.299
NSC0S11	10	33,605,721	33,948,896	32	0.316
NSC0S11	10	33,617,253	34,232,764	25	0.362
NSC0S11	10	33,750,281	34,232,921	30	0.325
NSC0S12	1	124,777,442	125,391,989	3	0.284
NSC0S12	2	49,929,930	50,455,279	7	0.214
NSC0S12	2	156,399,742	157,035,617	4	0.219
NSC0S12	3	53,146,873	53,741,054	2	0.243
NSC0S12	3	63,253,384	63,696,416	10	0.217
NSC0S12	4	135,481,568	136,002,833	16	0.239
NSC0S12	6	71,449,306	71,921,915	34	0.305
NSC0S12	6	71,449,535	72,008,063	23	0.435
NSC0S12	6	71,618,637	72,008,505	35	0.302
NSC0S12	6	71,620,957	72,179,048	35	0.312
NSC0S12	6	101,192,553	101,513,634	10	0.225

NSC0S12	7	113,535,012	114,178,113	3	0.370
NSC0S12	7	113,706,281	114,196,433	6	0.327
NSC0S12	8	36,689,480	37,228,414	14	0.216
NSC0S12	10	58,452,934	59,003,246	13	0.231
NSC0S12	10	58,581,071	59,044,804	13	0.220
NSC0S13	1	61,719,602	62,142,811	44	0.282
NSC0S13	3	89,608,459	90,147,946	6	0.352
NSC0S13	3	230,992,053	231,354,945	15	0.308
NSC0S13	4	36,418,456	36,914,505	66	0.318
NSC0S13	4	36,516,613	37,014,520	77	0.283
NSC0S13	4	69,997,680	70,631,467	5	0.362
NSC0S13	6	101,015,413	101,513,634	16	0.335
NSC0S13	6	101,192,502	101,801,332	16	0.317
NSC0S13	6	101,192,553	101,513,634	10	0.451
NSC0S13	6	101,492,170	101,991,919	20	0.281
NSC0S13	7	111,733,857	112,219,188	4	0.299
NSC0S13	9	27,334,668	27,951,991	23	0.272
NSC0S13	9	32,435,972	32,707,591	19	0.287
NSC0S13	9	32,583,788	33,125,927	19	0.289
NSC0S13	10	71,810,690	72,301,732	4	0.300
NSC0S13	10	73,205,989	73,748,093	25	0.288
NSC0S1	1	78,364,771	79,005,929	2	0.241
NSC0S1	3	89,608,459	90,147,946	6	0.241
NSC0S1	3	89,833,378	90,165,269	14	0.211
NSC0S1	3	90,033,872	90,417,914	19	0.199
NSC0S1	3	168,525,999	169,030,137	48	0.195
NSC0S1	4	76,857,174	77,360,921	9	0.192
NSC0S1	4	122,345,476	122,869,942	7	0.191
NSC0S1	4	140,349,938	140,878,179	4	0.266
NSC0S1	6	101,192,553	101,513,634	10	0.188
NSC0S1	7	26,599,643	27,032,384	20	0.214
NSC0S1	7	111,733,857	112,219,188	4	0.232
NSC0S1	7	113,535,012	114,178,113	3	0.324
NSC0S1	7	113,706,281	114,196,433	6	0.352
NSC0S1	9	64,411,571	64,808,356	18	0.210
NSC0S1	9	64,572,545	65,200,562	18	0.211
NSC0S1	10	82,235,650	82,721,188	19	0.194

6.3 Chapter 4 Supplementary Information

6.3.1 Supplementary Table 1.

List of 570 inbred lines, with genotype-by-sequence identifier (GBS ID), accession number (Accession), inbred name (Romay_Inbred), decade of release (Decade), era (Era), and genetic group (Group).

GBSID	Accession	Romay_Inbred	Decade	Era	Group
Ames19284:81N4HABXX:1:G6	Ames 19284	C103	1940	1	1
Ames19293:81FE7ABXX:5:A9	Ames 19293	Wf9	1930	1	1
Ames22756:81FE8ABXX:1:E6	Ames 22756	Os426	1930	1	1
Ames23393:81FENABXX:3:B9	Ames 23393	A96	1930	1	1
Ames23400:81FENABXX:7:D5	Ames 23400	A203	1940	1	1
Ames23401:81FE8ABXX:5:G7	Ames 23401	A204	1940	1	1
Ames23402:81FENABXX:4:E12	Ames 23402	A208	1940	1	1
Ames23403:81FE8ABXX:5:A10	Ames 23403	A218	1940	1	1
Ames23404:81N4HABXX:7:F7	Ames 23404	A223	1940	1	1
Ames23426:81FE8ABXX:5:C4	Ames 23426	A357	1940	1	1
Ames24589:81N4HABXX:7:H12	Ames 24589	I159	1930	1	1
Ames24590:81FE7ABXX:4:B3	Ames 24590	I198	1930	1	1
Ames26021:81FENABXX:2:E12	Ames 26021	P8	1930	1	1
Ames26028:81FE7ABXX:2:D5	Ames 26028	CI 540	1930	1	1
Ames26138:81FE8ABXX:7:H9	Ames 26138	CI 187-2	1930	1	1
Ames26604:81N4HABXX:5:A6	Ames 26604	38-11	1930	1	1
Ames26770:81FE7ABXX:1:E8	Ames 26770	66	1930	1	1
Ames26772:81FE7ABXX:3:H7	Ames 26772	H5	1940	1	1
Ames27066:81FE8ABXX:7:G8	Ames 27066	B164	1930	1	1
NSL30064:81FE7ABXX:4:C7	NSL 30064	W24	1940	1	1
NSL30071:81N4HABXX:7:E3	NSL 30071	W32	1940	1	1
NSL30904:81FENABXX:7:A1	NSL 30904	4226	1930	1	1
NSL32719:D0D0BACXX:3:A3	NSL 32719	ND5	1940	1	1
NSL32721:D0D0BACXX:3:H2	NSL 32721	ND36	1940	1	1
NSL32725:D0D0BACXX:3:G2	NSL 32725	ND211	1940	1	1
NSL65871:81FE7ABXX:4:C2	NSL 65871	I205	1930	1	1
NSL65874:81N4HABXX:6:H7	NSL 65874	Os420	1930	1	1
PI533659:81N4HABXX:4:A8	PI 533659	SD102	1940	1	1
PI538242:81N4HABXX:4:C7	PI 538242	SD106	1940	1	1
PI587146:81FE8ABXX:7:E3	PI 587146	Oh7	1940	1	1

R4:627C3AAXX:6:G11	R4	R4	1930	1	1
W22:8042MABXX:2:A7	W22	W22	1940	1	1
WD:627C3AAXX:5:G2	WD	WD	1930	1	1
A188:8042MABXX:3:D3	A188	A188	1940	1	2
Ames23389:81FE8ABXX:1:B10	Ames 23389	A15	1940	1	2
Ames23422:81FE8ABXX:5:D3	Ames 23422	A322	1940	1	2
Ames23424:81FE8ABXX:5:H1	Ames 23424	A340	1940	1	2
Ames23425:81FENABXX:7:B6	Ames 23425	A344	1940	1	2
Ames23429:81FENABXX:3:E7	Ames 23429	A385	1940	1	2
Ames26771:81FE8ABXX:2:D12	Ames 26771	33-16	1930	1	2
Ames27067:81FENABXX:7:F5	Ames 27067	C49A	1930	1	2
Ames27121:81FE7ABXX:2:F9	Ames 27121	K64	1940	1	2
Ames27122:81FE8ABXX:7:A2	Ames 27122	K148	1940	1	2
Ames27130:81FE8ABXX:2:D10	Ames 27130	Ky21	1930	1	2
Ames27136:81N4HABXX:2:D2	Ames 27136	Mo.G	1930	1	2
Ames28367:81FENABXX:5:D4	Ames 28367	CI 7	1930	1	2
K55:8042MABXX:2:B9	K55	K55	1940	1	2
NSL22634:81FE7ABXX:1:E5	NSL 22634	K155	1940	1	2
NSL30038:81FE8ABXX:7:C11	NSL 30038	K4	1930	1	2
NSL30056:81FE7ABXX:5:A4	NSL 30056	W703	1930	1	2
NSL30060:81FE7ABXX:3:F12	NSL 30060	W23	1930	1	2
NSL30065:81FE7ABXX:3:B8	NSL 30065	W9	1940	1	2
NSL32722:81FENABXX:7:H12	NSL 32722	ND255	1940	1	2
NSL32726:81FE8ABXX:3:G12	NSL 32726	ND167	1940	1	2
NSL32728:81FE8ABXX:4:D6	NSL 32728	ND230	1940	1	2
NSL32732:81FENABXX:6:F4	NSL 32732	ND203	1940	1	2
NSL32738:81FE8ABXX:4:B8	NSL 32738	ND283	1940	1	2
NSL42873:81FE8ABXX:1:E8	NSL 42873	A12	1940	1	2
NSL65873:81FE8ABXX:7:F3	NSL 65873	L317	1930	1	2
A131(Holland):D0D0BACXX:3:G5	A131	A131	1950	2	1
A239:627C3AAXX:5:B2	A239	A239	1950	2	1
A619:627C3AAXX:5:D3	A619	A619	1960	2	1
Ames19288:D0D0BACXX:3:G2	Ames 19288	Oh43	1950	2	1
Ames19326:81N4HABXX:5:B12	Ames 19326	R168	1960	2	1
Ames19330:81FENABXX:2:G8	Ames 19330	W182B	1950	2	1
Ames20140:D0D0BACXX:3:A1	Ames 20140	Mt42	1950	2	1
Ames22439:D0D0BACXX:3:H4	Ames 22439	A73	1950	2	1
Ames22767:81FE8ABXX:1:A5	Ames 22767	W59E	1950	2	1
Ames23391:81FENABXX:2:G9	Ames 23391	A34	1950	2	1

Ames23394:81FENABXX:3:H1	Ames 23394	A116	1950	2	1
Ames23397:81FENABXX:3:D6	Ames 23397	A158	1950	2	1
Ames23398:81FENABXX:2:E9	Ames 23398	A166	1950	2	1
Ames23406:81FE8ABXX:5:D7	Ames 23406	A251	1960	2	1
Ames23409:81FENABXX:3:A7	Ames 23409	A264	1960	2	1
Ames23410:81FENABXX:3:H2	Ames 23410	A265	1950	2	1
Ames23413:81FENABXX:7:E5	Ames 23413	A286	1950	2	1
Ames23415:81FENABXX:2:F11	Ames 23415	A295	1950	2	1
Ames23423:81FE8ABXX:5:F1	Ames 23423	A334	1950	2	1
Ames23427:81FENABXX:3:E2	Ames 23427	A374	1950	2	1
Ames23428:81FENABXX:3:D3	Ames 23428	A375	1950	2	1
Ames23430:81FE8ABXX:6:H7	Ames 23430	A401	1960	2	1
Ames23435:81FE8ABXX:5:G2	Ames 23435	A427	1950	2	1
Ames23440:81N4HABXX:4:A1	Ames 23440	A495	1960	2	1
Ames23442:81FENABXX:2:A11	Ames 23442	A498	1950	2	1
Ames23445:81FE8ABXX:4:H12	Ames 23445	A502	1960	2	1
Ames23446:81FE8ABXX:5:E10	Ames 23446	A508	1950	2	1
Ames23447:81FE8ABXX:6:F9	Ames 23447	A509	1950	2	1
Ames23448:81FE8ABXX:5:C8	Ames 23448	A547	1960	2	1
Ames23449:81FE8ABXX:6:D2	Ames 23449	A548	1960	2	1
Ames23450:81FE8ABXX:6:B10	Ames 23450	A556	1960	2	1
Ames23460:81FENABXX:2:F5	Ames 23460	A624	1960	2	1
Ames23475:81FE8ABXX:4:F12	Ames 23475	A637	1960	2	1
Ames23482:81FENABXX:7:H1	Ames 23482	A649	1960	2	1
Ames23483:81FE8ABXX:6:C10	Ames 23483	A650	1960	2	1
Ames23484:81FE8ABXX:6:F5	Ames 23484	A651	1960	2	1
Ames23922:81FE8ABXX:1:B7	Ames 23922	Oh51A	1950	2	1
Ames24706:81FE8ABXX:5:F12	Ames 24706	MS12	1960	2	1
Ames24707:81FENABXX:2:H9	Ames 24707	MS24	1960	2	1
Ames24708:81FE8ABXX:6:E8	Ames 24708	MS24A	1960	2	1
Ames24728:D0D0BACXX:3:G4	Ames 24728	MS107	1960	2	1
Ames24729:81N4HABXX:6:D2	Ames 24729	MS116	1960	2	1
Ames24730:81N4HABXX:3:E8	Ames 24730	MS132	1960	2	1
Ames24741:81FE8ABXX:5:D2	Ames 24741	MS206	1960	2	1
Ames24742:81FE8ABXX:5:F8	Ames 24742	MS211	1960	2	1
Ames24743:81FE8ABXX:6:D4	Ames 24743	MS214	1960	2	1
Ames24752:81FE8ABXX:6:A10	Ames 24752	MS1334	1960	2	1
Ames26116:81FENABXX:5:D7	Ames 26116	CI 3A	1950	2	1
Ames26774:81N4HABXX:6:A11	Ames 26774	H14	1950	2	1

Ames26808:81N4HABXX:2:A6	Ames 26808	Ill.Hy	1960	2	1
B2-good:627C3AAXX:7:E8	B2	B2	1950	2	1
CIze21:81FE7ABXX:2:F10	CIze 21	CI 21E	1950	2	1
NSL22629:81FE8ABXX:7:E9	NSL 22629	Oh7B	1950	2	1
NSL22630:81FENABXX:6:H1	NSL 22630	K150	1950	2	1
NSL22632:D0D0BACXX:4:F8	NSL 22632	K201	1950	2	1
NSL22635:81N4HABXX:1:H10	NSL 22635	K41	1950	2	1
NSL28965:81FE7ABXX:5:F2	NSL 28965	OH33	1950	2	1
NSL28966:81N4HABXX:6:F7	NSL 28966	Oh40B	1950	2	1
NSL29285:D0D0BACXX:3:B5	NSL 29285	R903B	1960	2	1
NSL29287:81FE7ABXX:6:A3	NSL 29287	R941	1960	2	1
NSL29305:81FE7ABXX:5:A10	NSL 29305	R218	1960	2	1
NSL29314:81FENABXX:1:F12	NSL 29314	R219	1960	2	1
NSL30067:81FE7ABXX:5:A1	NSL 30067	W22R	1950	2	1
NSL30073:81FE7ABXX:4:B6	NSL 30073	W37A	1950	2	1
NSL30861:81FE7ABXX:2:E8	NSL 30861	A	1960	2	1
NSL30863:81FE7ABXX:1:D1	NSL 30863	L	1960	2	1
NSL30865:D0D0BACXX:3:G11	NSL 30865	R2	1950	2	1
NSL30868:81FE8ABXX:4:E5	NSL 30868	R30	1960	2	1
NSL30869:81FE7ABXX:3:E8	NSL 30869	R53	1960	2	1
NSL30872:81FENABXX:1:F11	NSL 30872	R71	1960	2	1
NSL30873:81FENABXX:2:H4	NSL 30873	R74	1960	2	1
NSL30875:81FE7ABXX:5:D1	NSL 30875	R76	1960	2	1
NSL30878:81FE7ABXX:4:F4	NSL 30878	R101	1960	2	1
NSL30879:81FENABXX:1:G1	NSL 30879	R103	1960	2	1
NSL30880:81FE8ABXX:5:G8	NSL 30880	R105	1960	2	1
NSL30883:81FENABXX:2:B4	NSL 30883	R113	1960	2	1
NSL30885:81N4HABXX:7:C4	NSL 30885	R138	1960	2	1
NSL30888:81FE7ABXX:5:G10	NSL 30888	R154	1960	2	1
NSL30890:81N4HABXX:7:D1	NSL 30890	R159	1960	2	1
NSL30894:81FE7ABXX:4:G11	NSL 30894	R177	1950	2	1
NSL30895:81FENABXX:1:A5	NSL 30895	R181	1960	2	1
NSL30898:81FE7ABXX:4:A9	NSL 30898	R192	1960	2	1
NSL30901:81FENABXX:2:G5	NSL 30901	R196	1960	2	1
NSL30902:81FE7ABXX:3:A11	NSL 30902	R197	1960	2	1
NSL30906:81FE7ABXX:3:F8	NSL 30906	R904A	1960	2	1
NSL30907:81FENABXX:2:H7	NSL 30907	R906	1960	2	1
NSL30908:81N4HABXX:7:D9	NSL 30908	R907	1960	2	1
NSL30911:81FENABXX:1:B3	NSL 30911	R917A	1960	2	1

NSL30912:81FE7ABXX:5:E4	NSL 30912	R921E	1960	2	1
NSL32720:D0D0BACXX:3:D1	NSL 32720	ND33	1960	2	1
NSL32730:81FE8ABXX:4:A2	NSL 32730	ND405	1960	2	1
NSL42804:81N4HABXX:7:G11	NSL 42804	R223	1960	2	1
NSL42872:81FE8ABXX:7:G11	NSL 42872	A7	1950	2	1
NSL42875:81N4HABXX:6:A7	NSL 42875	A71	1950	2	1
NSL42877:81FE8ABXX:5:B7	NSL 42877	A165	1950	2	1
NSL42878:81FE8ABXX:6:F7	NSL 42878	A171	1950	2	1
NSL65864:81FENABXX:6:B3	NSL 65864	B8	1950	2	1
NSL65868:81FENABXX:2:C3	NSL 65868	B21	1950	2	1
NSL65869:D0D0BACXX:3:F7	NSL 65869	B38	1950	2	1
NSL81595:81FE8ABXX:6:G2	NSL 81595	A238	1960	2	1
NSL8581:81FE7ABXX:3:H3	NSL 8581	81-1	1960	2	1
NSL91618:D0D0BACXX:3:H8	NSL 91618	Oh28	1950	2	1
PI550518:81FE8ABXX:7:D8	PI 550518	T8	1950	2	1
PI587136:81FE8ABXX:7:E4	PI 587136	Ky228	1960	2	1
PI587154:81FE8ABXX:1:E1	PI 587154	W153R	1950	2	1
PI601009:D0D7RACXX:2:A2	PI 601009	B47	1950	2	1
R109B:8042MABXX:2:F1	R109B	R109B	1960	2	1
B37:627C3AAXX:6:B11	B37	B37	1950	3	2
A632:8042MABXX:2:F4	A632	A632	1960	2	2
A634:627C3AAXX:7:C1	A634	A634	1960	2	2
Ames19309:81FE7ABXX:4:E6	Ames 19309	A635	1960	2	2
Ames19310:81FE8ABXX:6:B1	Ames 19310	A639	1960	2	2
Ames19311:81FE8ABXX:6:B3	Ames 19311	A641	1960	2	2
Ames23474:81FE8ABXX:5:D4	Ames 23474	A636	1960	2	2
Ames23477:81FENABXX:4:D12	Ames 23477	A640	1960	2	2
Ames23488:81FE8ABXX:6:F6	Ames 23488	A656	1960	2	2
Ames24704:81FENABXX:2:H12	Ames 24704	MS1	1950	2	2
Ames24727:81FE8ABXX:5:C7	Ames 24727	MS106	1960	2	2
NSL29282:D0D7RACXX:1:A1	NSL 29282	R901	1960	2	2
NSL29303:81FE7ABXX:6:D3	NSL 29303	R216	1960	2	2
NSL29304:D0D0BACXX:4:G3	NSL 29304	R216A	1960	2	2
NSL30867:81FE8ABXX:1:H1	NSL 30867	M14	1960	2	2
NSL30910:81FE7ABXX:5:H12	NSL 30910	R914	1960	2	2
NSL30913:81FE7ABXX:5:C1	NSL 30913	R924	1960	2	2
NSL65866:81FE7ABXX:5:H4	NSL 65866	B14	1950	2	2
A554:627C3AAXX:5:C6	A554	A554	1960	2	3
Ames23465:81N4HABXX:2:F6	Ames 23465	A629	1960	2	3

Ames23466:81FE8ABXX:6:E4	Ames 23466	A630	1960	2	3
Ames23467:81FE8ABXX:6:G3	Ames 23467	A631	1960	2	3
Ames23476:81FE8ABXX:5:A2	Ames 23476	A638	1960	2	3
Ames23485:81FE8ABXX:5:A5	Ames 23485	A652	1960	2	3
Ames23486:81FENABXX:2:C11	Ames 23486	A653	1960	2	3
Ames23487:D0D0BACXX:3:F3	Ames 23487	A655	1960	2	3
H49:8042MABXX:3:A5	H49	H49	1950	2	3
NSL28967:81N4HABXX:5:H12	NSL 28967	Oh56A	1950	2	3
NSL29316:81N4HABXX:6:D7	NSL 29316	R220	1960	2	3
NSL29317:81FE8ABXX:1:B3	NSL 29317	R221	1960	2	3
NSL29319:81N4HABXX:6:A5	NSL 29319	R222	1960	2	3
NSL29898:81N4HABXX:1:G9	NSL 29898	R222A	1960	2	3
NSL30046:81FENABXX:5:H2	NSL 30046	B41	1950	2	3
NSL30874:81FENABXX:1:H1	NSL 30874	R75	1950	2	3
NSL30877:81FE7ABXX:5:H5	NSL 30877	R84	1950	2	3
NSL30889:81N4HABXX:7:A10	NSL 30889	R158	1960	2	3
NSL32739:81FENABXX:6:G3	NSL 32739	ND474	1960	2	3
PI587141:81FE8ABXX:4:G11	PI 587141	A654	1960	2	3
W64A:627C3AAXX:7:E9	W64A	W64A	1950	2	3
A661:627C3AAXX:7:G3	A661	A661	1970	3	1
Ames12725:81FENABXX:4:A10	Ames 12725	NC7	1970	3	1
Ames19002:81N4HABXX:4:G2	Ames 19002	Va17	1970	3	1
Ames20116:81FE8ABXX:3:G7	Ames 20116	Mo23W	1970	3	1
Ames20118:81N4HABXX:3:A5	Ames 20118	Mo25W	1970	3	1
Ames22442:81FE8ABXX:6:B7	Ames 22442	Ms100	1970	3	1
Ames23392:81FE8ABXX:5:B9	Ames 23392	A90	1970	3	1
Ames24705:81FE8ABXX:5:D1	Ames 24705	MS4	1970	3	1
Ames24713:81FENABXX:3:B4	Ames 24713	MS72	1970	3	1
Ames24734:81FENABXX:2:C12	Ames 24734	MS153	1970	3	1
Ames26775:81FENABXX:2:A8	Ames 26775	H19	1970	3	1
Ames26783:81N4HABXX:7:H1	Ames 26783	H41	1970	3	1
Ames26784:81FE7ABXX:6:G5	Ames 26784	H42	1970	3	1
Ames26785:81FENABXX:1:B6	Ames 26785	H45	1970	3	1
Ames26786:81FE7ABXX:4:C3	Ames 26786	H46	1970	3	1
Ames26791:81FE7ABXX:5:A6	Ames 26791	H55	1970	3	1
Ames26792:81FENABXX:1:H7	Ames 26792	H59	1970	3	1
Ames26794:81FENABXX:2:E6	Ames 26794	H73	1970	3	1
Ames27131:81FE8ABXX:7:D10	Ames 27131	Ky226	1970	3	1
Ames27137:81N4HABXX:7:B12	Ames 27137	N6	1970	3	1

Ames28927:81N4HABXX:3:A2	Ames 28927	Mo27W	1970	3	1
Ames28928:81FE7ABXX:7:D11	Ames 28928	Mo28W	1970	3	1
Ames28929:81FE8ABXX:3:C11	Ames 28929	Mo29W	1970	3	1
Ames28930:81FENABXX:5:B7	Ames 28930	Mo30W	1970	3	1
Ames28931:81FENABXX:3:G11	Ames 28931	Mo31W	1970	3	1
Ames28932:81FE7ABXX:1:G7	Ames 28932	Mo32W	1970	3	1
Ames28933:81FE7ABXX:1:G5	Ames 28933	Mo33W	1970	3	1
Ames28945:81FE7ABXX:2:C10	Ames 28945	Mo511W	1970	3	1
Ames28946:81FE7ABXX:2:C8	Ames 28946	Mo512AW	1970	3	1
Ames28947:81FENABXX:5:H12	Ames 28947	Mo512BW	1970	3	1
B75:627C3AAXX:7:D3	B75	B75	1970	3	1
B77:627C3AAXX:5:F10	B77	B77	1970	3	1
MS71:D0D7RACXX:2:H7	MS71	MS71	1970	3	1
NSL22631:D0D0BACXX:4:H5	NSL 22631	K11	1970	3	1
NSL30074:81FE8ABXX:5:F2	NSL 30074	W182E	1970	3	1
NSL30896:81FENABXX:1:E2	NSL 30896	R181B	1970	3	1
NSL32729:81FE8ABXX:4:D7	NSL 32729	ND468	1970	3	1
NSL32735:81FENABXX:6:D5	NSL 32735	ND407	1970	3	1
NSL32736:81FENABXX:6:A4	NSL 32736	ND480	1970	3	1
NSL34374:81FENABXX:7:G7	NSL 34374	SD15	1970	3	1
NSL65863:81FENABXX:1:A2	NSL 65863	B7	1970	3	1
NSL67786:81FENABXX:5:A5	NSL 67786	Mo301ae	1970	3	1
NSL67788:81FENABXX:4:H4	NSL 67788	Mo303ae	1970	3	1
NSL67789:81FENABXX:3:B12	NSL 67789	Mo304ae	1970	3	1
NSL67793:81FE7ABXX:2:H1	NSL 67793	Mo308ae	1970	3	1
NSL67796:81FENABXX:5:F2	NSL 67796	Mo311ae	1970	3	1
NSL67798:81FE7ABXX:1:C6	NSL 67798	Mo313ae	1970	3	1
NSL81596:81FENABXX:3:H6	NSL 81596	A258	1970	3	1
NSL81597:81FE8ABXX:6:E11	NSL 81597	A648	1970	3	1
NSL81598:81FENABXX:2:F12	NSL 81598	A657	1970	3	1
NSL81599:81FENABXX:7:H6	NSL 81599	A659	1970	3	1
NSL81600:81FENABXX:3:A4	NSL 81600	A660	1970	3	1
PI543916:81N4HABXX:3:B12	PI 543916	Mp339	1970	3	1
PI548792:81FE7ABXX:6:F9	PI 548792	SC213	1970	3	1
PI550442:81FE8ABXX:7:F11	PI 550442	Mo20W	1970	3	1
PI550443:81FE7ABXX:2:A1	PI 550443	B49	1970	3	1
PI550454:81FENABXX:4:C8	PI 550454	B52	1970	3	1
PI550463:81FE7ABXX:2:F5	PI 550463	B65	1970	3	1
PI550470:81FE7ABXX:5:C7	PI 550470	B54	1970	3	1

PI550471:81FE7ABXX:2:D8	PI 550471	B56		1970	3	1
PI550472:81FENABXX:4:F3	PI 550472	B57		1970	3	1
PI558516:81N4HABXX:1:C11	PI 558516	Mo7		1970	3	1
PI558517:81FE8ABXX:7:A9	PI 558517	Mo10		1970	3	1
PI558518:81N4HABXX:1:B8	PI 558518	Mo15W		1970	3	1
PI558520:81FE8ABXX:7:E8	PI 558520	Mo1W		1970	3	1
PI558521:81N4HABXX:2:E1	PI 558521	Mo2RF		1970	3	1
PI558525:81FE8ABXX:7:E10	PI 558525	Mo8W		1970	3	1
PI558526:81FE8ABXX:7:D12	PI 558526	Mo9W		1970	3	1
PI558527:81FENABXX:4:D6	PI 558527	Mo11		1970	3	1
PI558529:81FE8ABXX:7:G2	PI 558529	Mo13		1970	3	1
PI558531:81FE8ABXX:7:G6	PI 558531	Mo16W		1970	3	1
PI558533:81N4HABXX:3:F5	PI 558533	Mo21R		1970	3	1
PI587144:81FE8ABXX:2:G2	PI 587144	Mo24W		1970	3	1
PI587147:81FE7ABXX:1:D6	PI 587147	Pa91		1970	3	1
PI587148:81N4HABXX:3:F2	PI 587148	CI 66		1970	3	1
PI587150:D0D0BACXX:3:H2	PI 587150	Va35		1970	3	1
PI587153:81FE8ABXX:4:E8	PI 587153	W117		1970	3	1
PI601005:D0D0BACXX:3:E6	PI 601005		207	1970	3	1
PI603939:81FE7ABXX:3:B12	PI 603939	B42		1970	3	1
PI607519:D0D0BACXX:3:C1	PI 607519	ND240		1970	3	1
PI607522:D0D0BACXX:3:F2	PI 607522	A662		1970	3	1
PI607523:81FE8ABXX:6:D7	PI 607523	A663		1970	3	1
PI607527:81FE8ABXX:4:B4	PI 607527	ND300		1970	3	1
Ames19321:81FE7ABXX:4:G9	Ames 19321	N28		1970	3	2
Ames22748:81FE7ABXX:3:F10	Ames 22748	H100		1970	3	2
Ames26796:81FE7ABXX:4:G2	Ames 26796	H93		1970	3	2
Ames28934:81FENABXX:4:G11	Ames 28934	Mo36		1970	3	2
Ames28935:81FE7ABXX:1:E6	Ames 28935	Mo37		1970	3	2
Ames28936:81FENABXX:4:D11	Ames 28936	Mo38		1970	3	2
Ames28937:81FENABXX:4:F12	Ames 28937	Mo39		1970	3	2
B10:8042MABXX:2:G9	B10	B10		1970	3	2
B46:8042MABXX:4:C11	B46	B46		1970	3	2
B73:C07W2ACXX:1:C9	B73	B73		1970	3	2
B79:627C3AAXX:7:A6	B79	B79		1970	3	2
B84:627C3AAXX:5:A10	B84	B84		1970	3	2
H84:627C3AAXX:7:E6	H84	H84		1970	3	2
H91:627C3AAXX:5:C8	H91	H91		1970	3	2
PI550440:81FE7ABXX:4:D4	PI 550440	B64		1970	3	2

PI550455:81FENABXX:5:A11	PI 550455	B67	1970	3	2
PI550456:D0D0BACXX:3:A4	PI 550456	B69	1970	3	2
PI550461:D0D0BACXX:3:D12	PI 550461	B14A	1970	3	2
PI550462:D0D0BACXX:3:E4	PI 550462	B59	1970	3	2
PI550465:81FENABXX:4:E10	PI 550465	B68	1970	3	2
PI550468:81FENABXX:5:H6	PI 550468	B45	1970	3	2
PI550483:81FE7ABXX:5:E6	PI 550483	B76	1970	3	2
PI600772:D0D0BACXX:3:C5	PI 600772	FR19	1970	3	2
PI600958:81FE7ABXX:4:E1	PI 600958	FAPW	1970	3	2
PI607512:81FENABXX:2:E4	PI 607512	N7A	1970	3	2
PI607524:81FE8ABXX:3:F10	PI 607524	A664	1970	3	2
PI607525:81FENABXX:6:C7	PI 607525	A665	1970	3	2
Ames19016:81FE7ABXX:4:B11	Ames 19016	Va59	1970	3	3
Ames19313:81N4HABXX:7:D7	Ames 19313	C123	1970	3	3
Ames19319:81N4HABXX:7:G6	Ames 19319	H95	1970	3	3
Ames19324:81FE7ABXX:2:H6	Ames 19324	Pa871	1970	3	3
Ames26795:81FE7ABXX:6:G2	Ames 26795	H88	1970	3	3
Ames27184:81FE7ABXX:5:C6	Ames 27184	Pa762	1970	3	3
Ames27184:81FE7ABXX:5:C6	Ames 27184	Pa762	1970	3	3
Ames27193:81FE7ABXX:5:D2	Ames 27193	Va85	1970	3	3
NSL67792:81FE7ABXX:3:D5	NSL 67792	Mo307ae	1970	3	3
PI550464:81FE7ABXX:6:F1	PI 550464	B66	1970	3	3
PI550466:81FE7ABXX:3:E9	PI 550466	B70	1970	3	3
PI558524:81FE8ABXX:7:G3	PI 558524	Mo6	1970	3	3
PI558532:81FE7ABXX:6:D5	PI 558532	Mo17	1970	3	3
PI587129:81FENABXX:2:F9	PI 587129	H99	1970	3	3
PI587149:81FE8ABXX:1:H3	PI 587149	Va26	1970	3	3
PI600791:81N4HABXX:5:F12	PI 600791	LH38	1970	3	3
PI600944:81N4HABXX:1:C3	PI 600944	LH39	1970	3	3
PI608764:81FENABXX:4:C9	PI 608764	B55	1970	3	3
PI608777:81FE7ABXX:3:C3	PI 608777	B85	1970	3	3
PI608778:81FE7ABXX:1:E1	PI 608778	B86	1970	3	3
Ames26788:81FE7ABXX:4:A5	Ames 26788	H50	1970	3	4
Ames26789:81FE7ABXX:6:A1	Ames 26789	H51	1970	3	4
NSL67790:D0D7RACXX:1:E5	NSL 67790	Mo305ae	1970	3	4
NSL67799:81FE7ABXX:2:A3	NSL 67799	Mo314ae	1970	3	4
PI558522:81FE8ABXX:7:H12	PI 558522	Mo3	1970	3	4
PI550441:81FE8ABXX:7:B6	PI 550441	Mo18W	1970	3	5
PI558528:81N4HABXX:1:A9	PI 558528	Mo12	1970	3	5

PI558530:81N4HABXX:1:G11	PI 558530	Mo14W		1970	3	5
PI558534:81FE7ABXX:7:H12	PI 558534	Mo22		1970	3	5
Ames20119:81FENABXX:4:C11	Ames 20119	Mo40		1980	4	1
Ames20120:81FE7ABXX:2:D6	Ames 20120	Mo42		1980	4	1
Ames26801:81FENABXX:1:C9	Ames 26801	H108		1980	4	1
Ames26802:81FENABXX:1:C11	Ames 26802	H109		1980	4	1
Ames26909:81FE7ABXX:2:B8	Ames 26909	Mo41		1980	4	1
Ames28938:81FE7ABXX:1:C10	Ames 28938	Mo43		1980	4	1
PI547086-1:81FENABXX:7:H9	PI 547086	LH128		1980	4	1
PI600955:81FE7ABXX:4:E5	PI 600955	LH51		1980	4	1
PI601080:81FENABXX:1:G3	PI 601080	CR1HT		1980	4	1
PI601209:81FE7ABXX:4:D9	PI 601209	MBNA		1980	4	1
PI601316:81FE7ABXX:6:A7	PI 601316	LH54		1980	4	1
PI601317:81FENABXX:1:E5	PI 601317	LH57		1980	4	1
PI601360:D0D0BACXX:3:A5	PI 601360	LH52		1980	4	1
PI601416:81N4HABXX:7:A8	PI 601416	LH61		1980	4	1
PI601466:81FE8ABXX:3:E12	PI 601466	LH59		1980	4	1
PI601489:81FENABXX:1:D7	PI 601489		740	1980	4	1
PI601494:81N4HABXX:6:B8	PI 601494	LH65		1980	4	1
PI601612:81FE7ABXX:4:B12	PI 601612	S8326		1980	4	1
Ames22752:81FENABXX:1:G4	Ames 22752	H119		1980	4	2
Ames23496:81FENABXX:7:F6	Ames 23496	A672		1980	4	2
Ames27045:81FENABXX:6:H10	Ames 27045	CO220		1980	4	2
N192:627C3AAXX:7:D11	N192	N192		1980	4	2
NSL197105:81FENABXX:2:E1	NSL 197105	H117		1980	4	2
PI520771:81FENABXX:3:F1	PI 520771	W570		1980	4	2
PI537097:D0D0BACXX:3:H3	PI 537097	LH195		1980	4	2
PI537099:81FE7ABXX:4:G6	PI 537099	LH205		1980	4	2
PI538009:81FENABXX:1:H9	PI 538009	LH196		1980	4	2
PI539923:81FENABXX:2:G2	PI 539923	LH194		1980	4	2
PI539924:81N4HABXX:7:B2	PI 539924	LH202		1980	4	2
PI539927:D0D0BACXX:3:B4	PI 539927	LH193		1980	4	2
PI543847-2:81FENABXX:7:B8	PI 543847	PHV07		1980	4	2
PI543849-6:81FENABXX:7:D9	PI 543849	PHW51		1980	4	2
PI547088:D0D0BACXX:3:C5	PI 547088	LH208		1980	4	2
PI550527:81N4HABXX:4:H5	PI 550527	H111		1980	4	2
PI550565:81FE8ABXX:4:A7	PI 550565	N190		1980	4	2
PI587124:81FENABXX:3:F7	PI 587124	CM105		1980	4	2
PI587127:81FE7ABXX:5:G12	PI 587127	H105W		1980	4	2

PI600957:81FE8ABXX:5:F11	PI 600957	LH74	1980	4	2
PI600959:81FE8ABXX:3:A10	PI 600959	LH145	1980	4	2
PI600981:D0D0BACXX:3:C2	PI 600981	PHG39	1980	4	2
PI601004:81N4HABXX:6:B2	PI 601004	LH132	1980	4	2
PI601037:81FE7ABXX:6:B5	PI 601037	G80	1980	4	2
PI601159:81FE7ABXX:3:C12	PI 601159	6103	1980	4	2
PI601210:D0D0BACXX:3:G12	PI 601210	78004	1980	4	2
PI601211:81N4HABXX:7:A2	PI 601211	78010	1980	4	2
PI601301:81FE7ABXX:4:A1	PI 601301	78002A	1980	4	2
PI601362:81FE7ABXX:3:E1	PI 601362	PHW17	1980	4	2
PI601374:81FE8ABXX:5:C5	PI 601374	764	1980	4	2
PI601377:81FENABXX:1:H6	PI 601377	794	1980	4	2
PI601378:81FE7ABXX:5:G4	PI 601378	LP5	1980	4	2
PI601402:D0D0BACXX:3:A9	PI 601402	LH146Ht	1980	4	2
PI601417:81FE7ABXX:5:B12	PI 601417	NS701	1980	4	2
PI601439:81FE7ABXX:2:H2	PI 601439	FBHJ	1980	4	2
PI601441:81FE7ABXX:1:D3	PI 601441	PB80	1980	4	2
PI601442:81FENABXX:4:C2	PI 601442	PHG86	1980	4	2
PI601468:81FENABXX:2:D9	PI 601468	PHK29	1980	4	2
PI601491:81N4HABXX:6:C8	PI 601491	790	1980	4	2
PI601492:81FE8ABXX:5:G9	PI 601492	793	1980	4	2
PI601493:81FENABXX:5:G11	PI 601493	LH149	1980	4	2
PI601498:81FENABXX:1:G5	PI 601498	PHT55	1980	4	2
PI601500:81FE7ABXX:1:H10	PI 601500	PHV63	1980	4	2
PI601563:D0D0BACXX:3:E6	PI 601563	87916W	1980	4	2
PI601575:81FENABXX:5:C11	PI 601575	PHW52	1980	4	2
PI601610:81FE7ABXX:3:F6	PI 601610	H8431	1980	4	2
PI601611:81FE8ABXX:6:C1	PI 601611	S8324	1980	4	2
PI601729:81N4HABXX:7:E9	PI 601729	W8555	1980	4	2
PI601775:81FENABXX:1:E11	PI 601775	PHJ70	1980	4	2
PI601777:81FE8ABXX:1:A6	PI 601777	PHK35	1980	4	2
PI601808:81FE8ABXX:1:F5	PI 601808	2FACC	1980	4	2
Ames22751:81N4HABXX:7:H6	Ames 22751	H114	1980	4	3
Ames23492:81FENABXX:3:G4	Ames 23492	A668	1980	4	3
Ames23493:81N4HABXX:7:F11	Ames 23493	A669	1980	4	3
Ames23495:81FENABXX:7:F7	Ames 23495	A671	1980	4	3
Ames23498:81FENABXX:3:G3	Ames 23498	A674	1980	4	3
Ames23499:81FENABXX:3:C9	Ames 23499	A675	1980	4	3
Ames23500:81FE8ABXX:5:G6	Ames 23500	A676	1980	4	3

Ames27032:81FENABXX:7:F4	Ames 27032	CM145	1980	4	3
Ames28939:81FE7ABXX:5:H3	Ames 28939	Mo44	1980	4	3
NSL30835:81FENABXX:2:D11	NSL 30835	SD10	1980	4	3
Pa875:627C3AAXX:5:C10	Pa875	Pa875	1980	4	3
PI270297:81FE7ABXX:2:D12	PI 270297	B96	1980	4	3
PI508277:81FENABXX:3:F8	PI 508277	SD42	1980	4	3
PI509542:81FENABXX:5:G12	PI 509542	B89	1980	4	3
PI518660:81FENABXX:6:A9	PI 518660	SD43	1980	4	3
PI520772:81FENABXX:7:E7	PI 520772	W572	1980	4	3
PI531509:81N4HABXX:7:A6	PI 531509	R225	1980	4	3
PI531510:81N4HABXX:7:H10	PI 531510	R226	1980	4	3
PI531511:81FE7ABXX:6:C2	PI 531511	R227	1980	4	3
PI533658:D0D7RACXX:2:G4	PI 533658	SD101	1980	4	3
PI539920:81N4HABXX:7:D5	PI 539920	LH160	1980	4	3
PI543843:D0D0BACXX:3:B9	PI 543843	PHN34	1980	4	3
PI550496:81FE7ABXX:5:H7	PI 550496	H102	1980	4	3
PI550497:81FE7ABXX:4:A10	PI 550497	H103	1980	4	3
PI550504:81FE8ABXX:4:C3	PI 550504	ND248	1980	4	3
PI550526:81FENABXX:5:H3	PI 550526	H110	1980	4	3
PI550552:81FE8ABXX:4:A6	PI 550552	ND304W	1980	4	3
PI550554:81FENABXX:6:D7	PI 550554	ND256	1980	4	3
PI550556:81FENABXX:4:H2	PI 550556	B88	1980	4	3
PI550568:D0D0BACXX:3:C8	PI 550568	N194	1980	4	3
PI593460:81FENABXX:3:A9	PI 593460	991	1980	4	3
PI600956:81FE7ABXX:5:E10	PI 600956	MDF-13D	1980	4	3
PI601007:81FE8ABXX:4:A12	PI 601007	BO9	1980	4	3
PI601079:81FE7ABXX:4:H9	PI 601079	LH123HT	1980	4	3
PI601170:81FE8ABXX:6:D8	PI 601170	LH82	1980	4	3
PI601172:81FENABXX:5:D8	PI 601172	HBA1	1980	4	3
PI601230:81N4HABXX:1:F11	PI 601230	LH150	1980	4	3
PI601269:81N4HABXX:7:F9	PI 601269	5707	1980	4	3
PI601300:81FE8ABXX:1:B9	PI 601300	4676A	1980	4	3
PI601320:81FENABXX:4:C6	PI 601320	PHG84	1980	4	3
PI601321:81FE8ABXX:5:C2	PI 601321	PHJ40	1980	4	3
PI601322:81FENABXX:2:C5	PI 601322	PHZ51	1980	4	3
PI601361:81FENABXX:6:E3	PI 601361	PHR36	1980	4	3
PI601375:81FE8ABXX:1:D9	PI 601375	778	1980	4	3
PI601376:D0D0BACXX:3:A2	PI 601376	779	1980	4	3
PI601403:D0D0BACXX:3:F12	PI 601403	LH156	1980	4	3

PI601404:81N4HABXX:1:A8	PI 601404	LH60		1980	4	3
PI601430:C07W2ACXX:2:E2	PI 601430		807	1980	4	3
PI601438:81N4HABXX:6:A12	PI 601438	78371A		1980	4	3
PI601470:D0D0BACXX:3:E12	PI 601470	PHV78		1980	4	3
PI601490:D0D0BACXX:3:B8	PI 601490		787	1980	4	3
PI601496:81FE8ABXX:1:A10	PI 601496	PHK76		1980	4	3
PI601499:81FE7ABXX:2:E2	PI 601499	PHT77		1980	4	3
PI601501:81FE7ABXX:3:D9	PI 601501	PHW65		1980	4	3
PI601560:D0D0BACXX:3:B2	PI 601560	2MA22		1980	4	3
PI601562:81FENABXX:3:A12	PI 601562	78551S		1980	4	3
PI601568:D0D0BACXX:3:G6	PI 601568	PHM49		1980	4	3
PI601572:81FENABXX:1:B1	PI 601572	PHR47		1980	4	3
PI601576:81FE7ABXX:3:B3	PI 601576	PHW79		1980	4	3
PI601724:81FENABXX:2:G11	PI 601724	E8501		1980	4	3
PI601773:D0D0BACXX:3:H6	PI 601773	PHJ31		1980	4	3
PI601774:D0D0BACXX:3:A10	PI 601774	PHJ33		1980	4	3
PI601779:81FE7ABXX:4:G4	PI 601779	PHM57		1980	4	3
PI601781:81FE8ABXX:1:F8	PI 601781	PHN37		1980	4	3
PI601782:81N4HABXX:5:F11	PI 601782	PHN73		1980	4	3
PI601785:D0D0BACXX:3:G11	PI 601785	PHP60		1980	4	3
PI601789:81N4HABXX:5:D6	PI 601789	PHV37		1980	4	3
PI601790:81FE8ABXX:1:C7	PI 601790	PHW03		1980	4	3
PI601791:81FENABXX:4:F6	PI 601791	PHW20		1980	4	3
PI601792:81FENABXX:5:D11	PI 601792	PHW43		1980	4	3
PI606769:D0D7RACXX:1:G2	PI 606769	SD41		1980	4	3
PI644101:81FE7ABXX:2:G11	PI 644101	LH1		1980	4	3
SD44:627C3AAXX:6:G4	SD44	SD44		1980	4	3
PI524970:81FE8ABXX:5:G4	PI 524970	SD46		1980	4	4
PI543846:D0D0BACXX:3:B6	PI 543846	PHP76		1980	4	4
PI543850:D0D0BACXX:3:C9	PI 543850	PHW86		1980	4	4
PI559918:81FE7ABXX:5:D5	PI 559918	NQ508		1980	4	4
PI601006:D0D0BACXX:3:G10	PI 601006	PHG50		1980	4	4
PI601008:81FENABXX:1:F1	PI 601008	PHG35		1980	4	4
PI601150:81FE8ABXX:5:A4	PI 601150	PHG71		1980	4	4
PI601208:81FE8ABXX:5:E2	PI 601208	IB014		1980	4	4
PI601229:81FE7ABXX:5:G2	PI 601229	PHG83		1980	4	4
PI601270:81FENABXX:1:G9	PI 601270	PHG29		1980	4	4
PI601319:81FE8ABXX:5:F3	PI 601319	PHG72		1980	4	4
PI601457:81N4HABXX:6:H4	PI 601457	IB02		1980	4	4

PI601469:81FE7ABXX:2:G3	PI 601469	PHR25	1980	4	4
PI601495:81FENABXX:1:D9	PI 601495	PHK42	1980	4	4
PI601497:81FE7ABXX:1:B4	PI 601497	PHN11	1980	4	4
PI601558:81FENABXX:3:B6	PI 601558	11430	1980	4	4
PI601567:81FE7ABXX:4:H5	PI 601567	PHH93	1980	4	4
PI601569:81FENABXX:5:E12	PI 601569	PHN47	1980	4	4
PI601570:D0D0BACXX:3:F8	PI 601570	PHP02	1980	4	4
PI601583:81FE7ABXX:4:D1	PI 601583	NS501	1980	4	4
PI601725:81N4HABXX:6:G6	PI 601725	J8606	1980	4	4
PI601776:D0D0BACXX:3:H5	PI 601776	PHJ75	1980	4	4
PI601778:81FENABXX:3:D1	PI 601778	PHM10	1980	4	4
PI601780:81FE8ABXX:1:D4	PI 601780	PHN29	1980	4	4
PI601783:81FE7ABXX:6:G6	PI 601783	PHN82	1980	4	4
PI601786:81FE7ABXX:4:D6	PI 601786	PHR62	1980	4	4
PI601788:81FE7ABXX:2:D7	PI 601788	PHT22	1980	4	4
PI606768:81FENABXX:3:B11	PI 606768	SD40	1980	4	4
Ames19318:81FE7ABXX:4:G1	Ames 19318	H107	1980	4	5
Ames20190:81FE8ABXX:3:D12	Ames 20190	CM37	1980	4	5
Ames21512:81FENABXX:5:F7	Ames 21512	CI 31A	1980	4	5
Ames22749:81N4HABXX:7:C7	Ames 22749	H104W	1980	4	5
Ames22750:D0D7RACXX:1:C4	Ames 22750	H112	1980	4	5
Ames23490:81FE8ABXX:5:D8	Ames 23490	A666	1980	4	5
Ames23491:81FE8ABXX:6:D6	Ames 23491	A667	1980	4	5
Ames23494:81FENABXX:3:C1	Ames 23494	A670	1980	4	5
Ames23497:81FE8ABXX:5:H3	Ames 23497	A673	1980	4	5
Ames23501:81FE8ABXX:6:C2	Ames 23501	A677	1980	4	5
Ames23502:81FE8ABXX:6:H8	Ames 23502	A678	1980	4	5
Ames24714:81FENABXX:2:B10	Ames 24714	MS74	1980	4	5
Ames24737:81FE8ABXX:5:F6	Ames 24737	MS200	1980	4	5
Ames25559:81FE7ABXX:5:A5	Ames 25559	H60	1980	4	5
Ames26790:81FENABXX:1:B2	Ames 26790	H52	1980	4	5
Ames26793:81FENABXX:1:G12	Ames 26793	H71	1980	4	5
Ames26798:81FE7ABXX:3:H6	Ames 26798	H98	1980	4	5
Ames26799:81FE7ABXX:5:E11	Ames 26799	H101	1980	4	5
Ames26804:81FE7ABXX:5:A8	Ames 26804	H121	1980	4	5
Ames27070:81FE8ABXX:4:H5	Ames 27070	CM7	1980	4	5
Ames27105:81FE8ABXX:4:C8	Ames 27105	CO106	1980	4	5
Ames27139:81FE8ABXX:7:E11	Ames 27139	NC33	1980	4	5
Ames27194:81FE8ABXX:6:F3	Ames 27194	W401	1980	4	5

CMV3:627C3AAXX:7:A9	CMV3	CMV3	1980	4	5
ND246:8042MABXX:3:F6	ND246	ND246	1980	4	5
NSL197104:81FE7ABXX:5:H9	NSL 197104	H116	1980	4	5
NSL22983:81FENABXX:1:F9	NSL 22983	C102	1980	4	5
NSL42874:81FE8ABXX:1:A7	NSL 42874	C 14	1980	4	5
PI257506:81FE8ABXX:4:D1	PI 257506	F2	1980	4	5
PI520770:81FE8ABXX:5:A8	PI 520770	W552	1980	4	5
PI550489:81FE8ABXX:4:F3	PI 550489	ND245	1980	4	5
PI550491:81FENABXX:6:E6	PI 550491	ND301	1980	4	5
PI550503:81FENABXX:6:F12	PI 550503	ND247	1980	4	5
PI550530:81FE8ABXX:4:E3	PI 550530	ND249	1980	4	5
PI550531:81FENABXX:6:A12	PI 550531	ND250	1980	4	5
PI550532:81FE8ABXX:4:H3	PI 550532	ND251	1980	4	5
PI550540:81FE8ABXX:4:A8	PI 550540	ND252	1980	4	5
PI550541:81FE8ABXX:3:B9	PI 550541	ND253	1980	4	5
PI550542:81FENABXX:6:F7	PI 550542	ND254	1980	4	5
PI550553:81FE8ABXX:4:E1	PI 550553	ND101	1980	4	5
PI550561:81FENABXX:6:H12	PI 550561	ND257	1980	4	5
PI550567:81FENABXX:6:E10	PI 550567	N193	1980	4	5
PI550571:81FENABXX:6:D9	PI 550571	ND259	1980	4	5
PI550573:81FENABXX:6:G12	PI 550573	ND260	1980	4	5
PI583774:81FENABXX:3:E6	PI 583774	ML606	1980	4	5
PI587125:81FE8ABXX:3:B8	PI 587125	CO109	1980	4	5
PI601003:81FENABXX:6:H9	PI 601003	LH143 CMS	1980	4	5
PI601171:81N4HABXX:7:E8	PI 601171	LH93	1980	4	5
PI601318:81N4HABXX:7:C2	PI 601318	PHG47	1980	4	5
PI601405:81N4HABXX:7:G10	PI 601405	LH85	1980	4	5
PI601440:81FENABXX:1:H4	PI 601440	MBPM	1980	4	5
PI601467:81FENABXX:6:G8	PI 601467	PHK05	1980	4	5
PI601566:81FE7ABXX:3:C9	PI 601566	MBST	1980	4	5
PI601571:81FENABXX:4:G5	PI 601571	PHR32	1980	4	5
PI601574:81FENABXX:4:G9	PI 601574	PHT60	1980	4	5
PI608768:81FENABXX:3:E8	PI 608768	B87	1980	4	5